USCG

WORKSHOP

S
U
R
V
E
Y

The Design & Development of Survey Instruments

Performance Technology Center
Yorktown Virginia

By

James A. Pershing, Ph.D.
Education & Management Research Associates
Bloomington Indiana

26 September, 2000

Handbook

USCG

WORKSHOP

SURVEY

Table Of Contents

USCG

WORKSHOP

SURVEY

1 Questionnaire Development

2 Populations and Samples

3 Graphing Data

4 Descriptive Statistics

5 Inferential Statistics.

Section 1: Questionnaire Development

CONTENTS

Overview

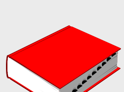
PAGE

Steps and Procedures

PAGE

Legend:

Reference notes and supplemental information on opposite page.

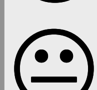


References to parts of the hand-book and other ideas.

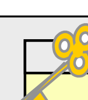
1 Numbered steps and Procedures

Important points, hints and ideas.

Advantages



Limitations



Key Points and Ideas

Figure 1.1: Examples of Closed-Form and Open-Form Questionnaires

Closed-Form Questionnaire:

Months With Company

- ☐ Less than one
- ☐ One to three
- ☐ Three to six
- ☐ Six to nine
- ☐ Nine to twelve

Job Title

- ☐ Operator
- ☐ Trouble-Shooter
- ☐ Maintenance Person

Shift

- ☐ Day
- ☐ Evening

Indicate your level of agreement with the following three statements:	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
1. I like my job.	SA	A	U	D	SD
2. I like my work shift.	SA	A	U	D	SD
3. I am willing to work overtime.	SA	A	U	D	SD

Open-Form Questionnaire:

1. List the jobs you have held during your tenure with the company. ______________________

2. Describe your present job. ______________________

3. Indicate what you like and dislike about your job.

4. How do you feel about working overtime? ______________________

Introduction

Written questionnaires are tools for collecting data by asking respondents questions or asking them to agree or disagree with statements representing different points of view. They are an indirect method of collecting data because they do not involve face-to-face interaction with respondents. Questionnaires give respondents time to think about their responses. They can be administrated relatively easily and inexpensively. When designed to do so, they can protect the anonymity of the respondents.

Written questionnaires are the most commonly used data gathering method in performance analysis. However, it is best to combine this method with one or two other methods.

Questionnaires are popular and versatile in performance analysis. In order to be used to measure employees' attitudes or feelings about performance and opportunities, items have to meet a high degree of applicability and technical quality (McBean & Al-Nassri, 1982). This means that the questions asked should be suitable to the intended purposes of the performance analysis project and show high levels of validity and reliability.

See a basic text on basic research techniques for more information related to validity and reliability.

Types of Questionnaires

There are two types of questionnaires: closed-form and open-form questionnaires. The closed-form provides a list of items to be checked, alternative answers to be selected, or blanks to be filled in. It can be fully structured with scales requiring "yes/no" or other types of directed responses. This type of questionnaire takes the respondent a minimum amount of time to complete, and tabulation is simple and not very time-consuming. However, the closed-form is difficult to design in that all possible responses to questions need to be anticipated in advance and included in the questionnaire.

Figure 1-1 provides examples of closed-form and open-form questionnaires.

The open-form of questionnaire offers the respondent an opportunity to give a more complete and comprehensive picture of a situation. It encourages respondents to convey their attitudes, feelings, opinions, and ideas in their own words. An open-ended questionnaire is less rigid and restrictive. However, it will take the respondent time to complete the questions. Tabulation and data analysis are difficult as well.

NOTES AND SUPPLEMENTAL INFORMATION

OVERVIEW

Figure 1.2: Task List for Developing Written Questionnaires

Plan for the Questionnaire

- ☐ Review advantages and limitations of using a questionnaire
- ☐ Prepare in writing objectives for the questionnaire
- ☐ Have objectives reviewed by others
- ☐ Prepare a work schedule and time-line for developing the questionnaire

Collect Information to Construct Questions

- ☐ Write items to collect demographic information
- ☐ Write items to collect factual information
- ☐ Write items to measure opinions and attitudes
- ☐ Write evaluative items

Conduct Item Try-Outs (Pre-Tests or Field Tests)

Prepare Draft Questionnaire

- ☐ Write descriptive title
- ☐ Write introduction
- ☐ Group items by content – subtitle each group
- ☐ Write instructions for returning the questionnaire
- ☐ If mailed questionnaire
  - Write cover letter
  - Write follow-up letter
- ☐ If group administered
  - Write instructions for administrator

Pre-Test Questionnaire

Revise and Administer

Questionnaire Elements

Figure 1.2 explains a task checklist for developing questionnaires.

There are five main elements to consider in designing and developing a good questionnaire: (1) introduction and closing; (2) directions; (3) question formats and rating scales; (4) question construction; and (5) layout or format.

Introduction and Closing

A questionnaire should have an introductory statement and a clear closing statement. The introductory statement presents the purpose of the questionnaire, asks for cooperation, and explains confidentiality or anonymity procedures.

The closing statement thanks respondents for completing the questionnaire and instructs them on returning the completed questionnaire. This guideline is related to logistics. Questionnaires should include what to do after completing the answers (Dixon, 1990). An example is: “When you have completed the questionnaire, please return it in the envelope included with the questionnaire.”

Directions

A self-administered questionnaire should begin with general instructions to be followed for completing it. For closed-ended questions such as multiple-choice, yes-no, and rating scales, respondents should be given instructions about using the answer formats such as placing a check mark or an “X” in the box beside the appropriate answer or by writing in their answers when called for. For open-ended questions such as fill-ins, short-answers, and essays, respondents should be given guidance as to whether brief or lengthy answers are expected.

Question Format

Closed-form Questions are structured questions.

O**pen-form Questions** are unstructured or semi-structured questions.

Questions for performance analysis fall into two general categories: closed-ended (structured) and open-ended (unstructured) questions. In closed-ended questions respondents are asked to select their answer from a fixed set of response alternatives. Closed-form questions are commonly used in needs analysis because of a greater uniformity of responses and easy administration. Their main drawback is in the structuring of responses (Babbie, 1990; Edwards, Thomas, Rosenfeld, & Booth-Kewley, 1997).

Open-ended questions ask respondents to provide their own answers to the questions. They give respondents a chance to answer using their own frame of reference, without undue influence from prefixed alternatives. However, the responses must be coded prior to data analysis, and there is a chance that some respondents will give irrelevant answers in terms of the purposes of the performance analysis.

NOTES AND SUPPLEMENTAL INFORMATION

OVERVIEW

Figure 1.3: Examples of Likert and Semantic Differential Scales

Likert Scale Example

	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
1. My immediate supervisor treats me fairly.	SA	A	U	D	SD
2. My immediate supervisor works hard.	SA	A	U	D	SD
3. The company's reengineering effort is worthwhile.	SA	A	U	D	SD
3. The company is sincere about implementing its reengineering effort.	SA	A	U	D	SD

Semantic Differential Scale Examples

My immediate supervisor is:

Fair	____	____	____	____	____	____	____	Unfair
Active	____	____	____	____	____	____	____	Passive
Hardworking	____	____	____	____	____	____	____	Lazy
Efficient	____	____	____	____	____	____	____	Inefficient

The concept of reengineering is:

Valuable	____	____	____	____	____	____	____	Worthless
Sincere	____	____	____	____	____	____	____	Insincere
Strong	____	____	____	____	____	____	____	Weak
Relaxed	____	____	____	____	____	____	____	Tense
Active	____	____	____	____	____	____	____	Passive
Warm	____	____	____	____	____	____	____	Cold
Fast	____	____	____	____	____	____	____	Slow

OVERVIEW

Rating Scales

A rating scale yields "a *single score* that indicates both the direction and intensity of a person's attitude" (Henerson, Morris, & Fitz-Gibbon, 1978, p. 84). Because the scoring method of most rating scales is based on the idea of measuring the intensity, hardness, or potency of a response (Neuman, 1997), each item must differentiate between those respondents with a favorable attitude from those with an unfavorable attitude. In addition, the items must allow for expression of a broad range of feelings from strongly favorable through neutral to strongly unfavorable. There are two rating scales commonly used in designing performance analysis questionnaires: Likert scale and Semantic Differential scale.

The Likert scale is the most commonly used scale in performance analysis questionnaires.

Likert Scale**:** Likert's (1932) scales, called summated-rating or additive scales, are widely used and very common because of easy construction, high reliability, and successful adaptation to measure many types of affective characteristics (Nunnally, 1978). On the Likert rating scale, a respondent indicates his/her agreement or disagreement with a variety of statements on an intensity scale. The 5-point strongly agree to strongly disagree format is commonly employed. Responses are then summed across the items to generate a score on the affective instrument.

The simplicity and ease of use of the Likert scale is its real strength. The Likert scale can provide an ordinal-level measure of a person's attitude (Babbie, 1995). The rating scales have the advantage of providing data that use values rather than merely categories (Edwards, Thomas, Rosenfeld, & Booth-Kewley, 1997). This can provide greater flexibility for data analysis.

Figure 1.3 provides examples of questions using Likert and Semantic Differential scales.

Semantic Differential Scale: Osgood's Semantic Differential scale provides an indirect measure of how a person feels about a concept, object, or other person (Neuman, 1997). The scale measures subjective feelings toward something by using a set of scales anchored at their extreme points by words of opposite meaning (Edwards, Thomas, Rosenfeld, & Booth-Kewley, 1997).

In order to use the Semantic Differential scale, a researcher presents target subjects with a list of paired opposite adjectives on a continuum of 5 to 11 points. Respondents mark the place on the scale continuum between the adjectives that best expresses their perceptions, attitudes, feelings, and the like. The results of Semantic Differential scales can be used to assess respondents' overall perceptions of different concepts or issues.

Studies of a wide variety of adjectives in English found that they fall into three major classes of meaning: evaluation (good-bad), potency (strong-weak), and activity (active-passive) (Neuman, 1997). Of the three classes of meaning, evaluation is usually the most significant. Semantic Differential scales yield interval data that are usable with virtually any statistical analysis. However, it is often difficult to give concise written directions for semantic differentials, especially to respondents unfamiliar with the rating scale.

NOTES AND SUPPLEMENTAL INFORMATION

OVERVIEW

Figure 1.4: Three Conditions for a Good Question

Kent (1993) provides **three conditions** to maximize valid responses to items on a questionnaire.

- Respondents must understand the questions, and understand them in the same way as other respondents.
- Respondents must be able to provide the answers.
- Respondents must be willing to provide the information. (p.78)

Figure 1.5: Examples of Common Wording Problems and Possible Solutions

Wording Problem:

Vaguely worded question and responses

Do you actively support, support, or not support use of quality teams in departments?

_____ Actively Support
_____ Support
_____ Not Support

Possible Revision:

A proposal has been made to expand the use of quality teams in your department. Would you be willing or not willing to participate 30 minutes per day in a quality team?

_____ Willing
_____ Not Willing
_____ Undecided

KEY POINT: Vaguely worded questions and responses produce questionable information. The problem question does not provide respondents information they can act on because they cannot tell what "support" means.

Wording Problem:

Use of jargon and acronyms respondents may not understand.

Do you believe IT programs offered by T&D should or should not be funded by operational units?

_____ Should
_____ Should Not
_____ Undecided or Unsure

Possible Revision:

The Finance Office is considering directly charging departments when they send personnel to company-sponsored education and training programs. Do you favor or not favor such a change?

_____ Favor
_____ Do Not Favor
_____ Undecided or Unsure

KEY POINT: Jargon and undefined acronyms should not be used unless the population being studied uses the terms regularly and will not find them confusing.

Wording Problem:

Too much precision.

How many times, if any, were you, yourself, delayed in your work because of an equipment failure last month? ____________________

Possible Revision:

How many times, last month, was your work interrupted because of an equipment failure?

_____ Never
_____ 1 - 4
_____ 5 - 8
___ 9-12
___ 13-20
___ More than 20

KEY POINT: Sometimes, too much precision makes a question impossible to answer for the respondents. It may provide more detail than the performance analysts actually needs. Categorical responses are easier for respondents.

OVERVIEW

Question Construction

Figure 1.4 offers three conditions for a good question.

Figure 1.5 provides guidelines for question writing.

Writing a question is more of an art than a science. It takes skill, practice, patience, and creativity. Neuman (1997) suggests two main principles for questions: "avoid confusion and keep the respondent's perspective in mind" (p. 233). Following are general guidelines for question writing and hints to avoid common errors.

- ☐ ***Write simple, clear, and short questions***: In order to avoid ambiguity and confusion, questionnaire items should be simple, clear, and kept as short as possible. The longer the question, the more difficult the task of answering. Although there is no magic number of words, Payne (1951) insists on using 25 or fewer words for a question.

- ☐ ***Make specific and precise questions***: Specific questions are usually better than general questions because they get similar interpretation by all of the respondents. Questions with specific and concrete wording are more apt to communicate the same meaning. Avoid wording such as frequently, most, sometimes, or regularly (Dixon, 1990).

- ☐ ***Use appropriate language***: Questions should be worded at an appropriate level for the respondents. Professional jargon, slang, technical terms, and abbreviations can carry different meanings to the respondents who vary in life and work experiences as well as education (Edwards, Thomas, Rosenfeld, & Booth-Kewley, 1997; Neuman, 1997). Avoid such terms unless a specialized population is being used as respondents.

- ☐ ***Ensure respondents' capabilities to answer*****:** In making questions, we should continually ask ourselves whether the respondents are able to provide reliable answers (Babbie, 1995). Asking questions that only a few respondents can answer frustrates most respondents and results in poor-quality responses. Asking the respondents to recall past details, answer specific factual information, or make a choice about something they know nothing about may result in an answer, but one that is meaningless (Neuman, 1997).

- ☐ ***Ask only one topic or idea per item***: Each question should be related to only one topic or idea. Question items containing two ideas or those which combine two questions into one are called 'double-barreled' (Babbie, 1995; Neuman, 1997). The problem with a double-barreled question is that agreement or disagreement with the item implies agreement or disagreement with both parts of it. The best way of dealing with double-barreled questions is to break them up and list one question per idea or topic.

- ☐ ***Uses of closed-ended and open-ended questions***: Closed-ended questions take longer to develop, obtain a single answer or choice from several options, take a shorter time to complete, and are more easily quantified for data analysis. While open-ended questions provide in-depth responses and unanticipated information, take longer to be complete, and take longer to analyze. Generally, use both types of questions in a questionnaire.

- ☐ ***Use appropriate emphasis for key words in the question***: The use of an emphasis tool such as boldface, italicized, capitalized, or underlined words or phrases within a question can help to clarify potential confusion (Rea & Parker, 1992).

Figure 1.5: Examples of Common Wording Problems and Possible Solutions (continued)

Wording Problem:

Double-barreled questions.

Enough people and equipment exist to complete scheduled maintenance activities.

_____ Agree _____ Unsure _____ Disagree

Possible Revision:

Enough people exist to complete scheduled maintenance activities.

_____ Agree _____ Unsure _____ Disagree

The necessary equipment is in place to complete scheduled maintenance activities.

_____ Agree _____ Unsure _____ Disagree

KEY POINT: Double-barreled questions are questions with multiple parts. They use words such as 'and' and 'or' and use commas or dashes to separate listed parts. Sometimes, 'etc.' is used. Such questions produce ambiguous answers because we do not know if the respondent is referencing all or only parts of the question. Split the question into parts so respondents can answer one part at a time.

Wording Problem:

Negative Items.

Manufacturing does not measure the costs of its activities.

_____ Yes
_____ No
_____ Do Not Know

Possible Revision:

Manufacturing measures the costs of its activities.

_____ Yes
_____ No
_____ Do Not Know

KEY POINT: The appearance of a negation in an item leads to easy misinterpretation. A sizeable number of respondents will read over the 'not' word and answer on that basis. Others will read and consider the negative word. The needs analysis will not know which is which. Change negatively worded items to positive wording.

Wording Problem:

Bias from slanted introduction.

Our CEO believes that healthy workers are more productive workers. Do you exercise - such as bike, walk, or swim - regularly, or do you not exercise regularly?

_____ Do Exercise Regularly
_____ Do Not Exercise Regularly

Possible Revision:

Do you exercise - such as bike, walk, or swim -regularly, or do you not exercise regularly?

_____ Do Exercise Regularly
_____ Do Not Exercise Regularly

KEY POINT: Bias is created when a question is associated with a person of note, when the 'tone' is subjective, or when the question makes it seem 'everyone' has a belief or engages in a particular activity. Eliminate bias from questions.

☐ ***Take care with sensitive questions***: Asking sensitive questions has always been a difficult issue (Edwards, Thomas, Rosenfeld, & Booth-Kewley, 1997). People vary in the amount and type of information they are willing to disclose about items such as their salary, race, or ethnic status. In dealing with these kinds of sensitive questions, special care should be taken. It is also necessary to be careful when using terms of regional, occupational, or social class difference.

☐ ***Avoid negative questions and double negatives***: The word ‘not’ in a negative question can easily pave the way for misinterpretation. Questions with double negatives are confusing and difficult. A double-negative question may ask respondents to disagree that something in a question is false or negative. This results in “ an awkward statement and a potential source of considerable error” (Sheatsley, 1983, p. 217).

Figure 1.5 suggests guidelines for question writing.

☐ ***Avoid biased or loaded questions and terms*****:** The way in which questions are worded, or the inclusion of certain terms, may encourage some respondents more than others. Such questions are called ‘biased or loaded’ and should be avoided in questionnaires (Babbie, 1995; Neuman, 1997). There are many sources to bias a question, such as identification of a well-known person, agency, or social desirability.

☐ ***Avoid questions with false premises or future intentions*****:** Respondents who disagree with the premises will be frustrated in answering. If it is necessary to include questions with false premises, the question should explicitly ask the respondents to assume the premise is true, then ask for a preference.

☐ ***Avoid overlapping or unbalanced response categories***: Response categories should be mutually exclusive, exhaustive, and balanced (Neuman, 1997). In order to avoid overlapping responses, response categories should be mutually exclusive. Exhaustive means that every respondent has one choice or way to answer a question. Response categories should be kept balanced – that is, with an equal number of positive and negative options for the respondent to choose from.

Layout and Format

Prepare questionnaire to have a professional appearance. Use word-processing or desktop-publishing software.

The appearance and arrangement of the questionnaire should be clear, neat, and easy to follow. Often respondents seem to decide whether they will participate based on the appearance of the questionnaire (Moran, 1990). A professional appearance with high-quality graphics, space between questions, and good layout improves accuracy and completeness and helps the questionnaire flow (Kent, 1993; Neuman, 1997). Squeezing as many questions as possible on a page makes the questionnaire shorter in pages, but the clutter may result in overlooked questions or in respondents deciding not to participate.

A basic requirement for questionnaire layout is to give respondents adequate instructions for completing the questions. Instructions for items need to be distinguishable from the questions themselves and from the pre-coded answers. Instructions can be underlined, bold, or in a different or larger font. Often, response items are capitalized and the questions are written using lower case letters (Kent, 1993).

Figure 1.6: Advantages and Limitations of Questionnaires

Advantages of Questionnaires

- A questionnaire can provide insight into the perceptions of people in an organization. It is a suitable method for collecting data about known problems or clearly understood opportunities. It is a sensitive barometer of organizational culture.
- The method can be confidential or anonymous and respondents may freely give information without fear or reprisal. Questions can be answered at the respondent's convenience and personal pace, without probing by an interviewer. More people can be reached with the written questionnaire than any other type of survey.
- The written questionnaire is usually the least expensive type of survey. The method can elicit information from a wide number of respondents in a relatively short period of time. No specialized training is necessary to administrate the questionnaire or tabulate results.
- The questionnaire must be structured in advance in order to facilitate the processing of the results. Since all respondents receive the exact same questions in written form, answers are not susceptible to the biases that can slip into personal interviews.
- Results are easily quantifiable. For example, questionnaires’ data can be analyzed with little difficulty.
- Many standard questionnaires are available, eliminating development time and effort. Needs analysts can use them as is or add to and modify them.

Limitations of Questionnaires

- Excellent questionnaires, especially complex ones, are difficult to design. Development of effective and reliable instruments which are complex can be costly and requires strong technical skills.
- Questionnaires are influenced by the level of ambiguity in both questions and answers. Since communication is one way, respondents may not properly interpret questions and their answers can also be misinterpreted. If the questionnaire comes by mail, there is no facilitator to clarify questions about items on the questionnaire.
- Low return rates and inappropriate responses hinder accuracy. Respondents often are irritated by ’yet another survey to answer,' so response rate or accuracy may be poor.
- Results may be misleading, because only those respondents particularly interested in the content or outcome may respond. Such problems are multiplied with open-ended questions.
- Because considerable subjectivity comes into play in interpreting answers to open-ended questions, the data can be inefficient and subjective when used as the only source of information.

OVERVIEW

Layout is crucial for self-completed or mail questionnaires because there is no friendly personnel to interact with the respondent. Instead, the questionnaire's appearance persuades the respondent's answers.

Mail questionnaires should include a polite, professional cover letter on letterhead stationery, identification of the needs analyst, telephone or facsimile numbers to use for questions, and a statement of appreciation for participation (Neuman, 1997). Questionnaires should leave respondents with a positive feeling about the performance analysis and a sense that their participation is both needed and appreciated.

Figure 1.6 offers Questionnaire

☺ Advantages
☹ Limitations

Advantages and Limitations of Questionnaires

Planning and constructing written questionnaires can be a simple or a difficult task. For complex questionnaires, it can be time-consuming and expensive. Many steps are involved and some require specialized knowledge and skill.

Further Readings

Babbie, E. (1995). The practice of social research (7th ed.). Belmont, CA: Wadsworth.

Edwards, J. E., Thomas, M. D., Rosenfeld, P., & Booth-Kewley, S. (1997). How to conduct organizational surveys: A step-by-step guide. Thousand Oaks, CA; Sage.

Jones. J.E. & Bearley, W.L. (1995). Surveying Employees: A practical guidebook. Amherst, MA' HRD Press.

Patten, M.L. (1998), Questionnaire research: A practical Guide. Los Angeles, CA; Pyrczak.

NOTES AND SUPPLEMENTAL INFORMATION

STEPS AND PROCEDURES

Table 1.1: Comparative Survey Techniques

FEATURE	WRITTEN QUESTIONNAIRE	TELEPHONE INTERVIEW	FACE-TO-FACE INTERVIEW
COST	Inexpensive if kept simple No interviewers Printing and mailing costs	Moderate Interviewer time No transportation	Expensive Interviewer time, travel, and transportation
RESPONSE RATE	Low without incentives or extensive follow-up	Good 60% or better	High 75% or better
DATA COLLECTION TIME	Four to six weeks	One or two weeks	Two or three weeks
QUALITY OF ANSWERS	Good candor Cannot probe Can be anonymous	Succinct answers Can be confidential	In-depth responses Can be confidential
EXPERTISE TO CARRY OUT	High expertise on questionnaire development None on interviewing skills	Minimal expertise Moderate degree of interviewer training	High expertise High degree of interviewer training
RESPONDENT TIME TAKEN	Short 10-15 minutes typically	Moderate 15-30 minutes	Long 45-60 minutes

Adapted from: Weisberg, Herbert F., Krosnick, John A., and Bowen, Bruce D. **An Introduction to Survey Research, Polling, and Data Analysis, 3rd Edition**. Thousand Oaks: Sage Publications, 1996, pg. 121.

STEPS AND PROCEDURES

Using the Written Questionnaire for Performance Analysis

- DO NOT use a written questionnaire as your only source of needs analysis data.
- It is best to use a questionnaire AFTER using more exploratory methods such as observations, interviews, focus groups, or document analysis.
- Familiarize yourself with the advantages and disadvantages of the written questionnaire.
- BEST USES of written questionnaires ask for the following types of information from individuals:
  - Informed opinions based on personal experience and expertise
  - Knowledge or facts about themselves
  - Knowledge or facts about others for which they have direct knowledge

See page 1-12 to review the advantage and limitations of questionnaires.

Table 1.1 presents comparisons of written questionnaires to similar survey methods.

USE THE FOLLOWING TEN STEPS TO DEVELOP AND ADMINISTER A QUESTIONNAIRE

1 Write a Statement of Purpose

1-1 Write down the answers to the following 4 questions.

- Who are the champions for the perceived performance improvement initiative?
- What performance problem, quality improvement initiative, or business opportunity is to be addressed by the questionnaire?
- What objectives are to accomplished by the questionnaire?
- What will be done with the questionnaire results?

1-2 Using the answers to the above 4 questions, write a brief statement about what the questionnaire is to accomplish.

1-3 Share the paragraph with:

- Colleagues working on the needs analysis project
- The champions of the performance analysis initiative

1-4 If there are concerns or disagreements, reconcile the differences and:

- Re-write the statement of purpose, OR
- Settle upon an alternative data collection method

Prepare a written list of specific objectives for the questionnaire.

NOTES AND SUPPLEMENTAL INFORMATION

STEPS AND PROCEDURES

Table 1.2 Using Archival Data to Identify Content for a Questionnaire

PERFORMANCE NEED ADDRESSED IN A QUESTIONNAIRE (Examples)	POTENTIALLY RELEVANT ARCHIVAL DATA (Examples)
TASK INVENTORY	■ Existing job descriptions ■ Training manuals ■ Production delays
SAFETY IMPROVEMENT	■ Injury reports ■ Safety training requirements
ABSENTEEISM	■ Personnel records ■ Company policies on absenteeism

Adapted from: Edwards, Jack E., Thomas, Marie D., Resenfeld, Paul, and Boothe-Kewley, Stephanie. **How to Conduct Organizational Surveys: A Step-by-Step Guide.** Thousand Oaks: Sage Publications, 1997, pg. 21.

Figure 1.7: Attributes of Closed-Ended and Open-Ended Questions

CLOSED-ENDED QUESTIONS

- Structured items which give response alternative from which the respondent MUST choose
- Fast and easy for respondent
- Easy to code and process responses
- Same frame of reference for all respondents
- Provides respondent with cues to retrieve information from memory
- May force respondent to choose among alternatives which do not reflect their real judgments or feelings about the topic
- May get responses when people have no opinion on a matter or they do not understand the question
- May not include all aspects of the topic under consideration

OPEN-ENDED QUESTIONS

- Unstructured items which require respondents to supply answers to questions using their own words
- Takes time and thought on the part of the respondent
- Responses must be read and carefully analyzed to accurately code and summarize the information
- Obtain a more in-depth understanding of respondents' opinions and attitudes
- Allows respondents to vent their feelings about emotion-laden topics
- Respondents sometimes appreciate the opportunity to express their feelings
- Champions and managers interested in the questionnaire results often like to hear the opinions of respondents in their "own words"
- If long narratives are required, they take a long time to complete and analyze
- Assumes respondent is capable and comfortable writing narrative answers
- Not effective in assessing intensity of feelings or opinions about a topic

STEPS AND PROCEDURES

2 Identify Questionnaire Content

Table 1.2 explains uses of archival data that are often overlooked in performance analysis

2-1 Identify sources of data to use in developing the content of the questionnaire. Typical sources include:
- Interviews
- Focus groups
- Observations
- Past surveys
- Professional literature
- Archival sources

2-2 Cluster the raw information collected into survey categories, items, and response categories.
- Use the "cut-and-paste method" to put together information in similar categories
- Use an "Affinity Diagram" method to place the information into clusters or categories

See a basic TQM Book for Affinity Diagram guidelines.

3 Write Questionnaire Items

3-1 For each cluster or category of information to be asked on the questionnaire, decide what type of information is needed from the respondents. You can ask people:
- What they do – their behavior
- About themselves – their attributes
- What they think is true – their beliefs or perceptions
- What they want – their attitudes

Figure 1.7 illustrates attributes of closed-ended and open-ended questions.

3-2 Decide upon the question structure – open- or closed-ended questions. In making the choice, consider the following:
- Closed-ended questions are preferable to assess intensity -- that is, the strength of feelings a respondent has towards an issue or topic
- Use open-ended questions when enough is not known about a topic to write a comprehensive and inclusive list of alternative responses
- Allow room for comments at the end of a section of closed-ended questions so that respondents can add relevant information in their own words
- Provide open-ended questions at the end of a questionnaire asking for general comments, or other comments about topics covered and not covered in the questionnaire
- Most performance analysis questionnaires use a combination of closed- and open-ended questions

Figure 1.8: A Questionnaire Item Structured Four Different Ways

Open-Ended:

In your judgment, what problems face purchasing departments in the next five years?

____

____

____

Close-Ended -- Ordered Responses:

Listed below are three problems some people believe purchasing departments will face in the next five years. In your judgment, how serious is each one? (Circle your answer.)

a) Hiring qualified workers	Very	Somewhat	Not At All
b) Keeping up with technology changes	Very	Somewhat	Not At All
c) Finding reliable suppliers	Very	Somewhat	Not At All

Close-Ended -- Unordered Responses:

In your judgment, which one of the following problems facing purchasing departments in the next five years is the **MOST** serious? (Circle the number for your response.)

1 Hiring qualified workers
2 Keeping up with technology changes
3 Finding reliable suppliers

Partially Close-Ended:

In your judgment, which one of the following problems facing purchasing departments in the next five years is the **MOST** serious? (Circle the number for your response.)

1 Hiring qualified workers
2 Keeping up with technology changes
3 Finding reliable suppliers
4 Other (Please Specify) ________________________________

STEPS AND PROCEDURES

3-3 Format open-ended questions using the following guidelines.

- Make the items clear and distinct
- Be precise and make certain the respondent knows exactly what it is you are asking
- Provide enough space for the respondent to answer the question

3-4 Format closed-ended questions using the following guidelines.

- **Yes-No** or **True-False** formats
  - Primary use is to gather factual data – Example: Have you been late for work in the past four weeks?
  - Simple to use and analysis is straightforward
  - NOT appropriate for complex issues
  - CANNOT assess differences in intensity
  - Can add third category such as "Not applicable" or "Do not know"
- **Multiple Choice** formats
  - Choose one or more answers from a list of alternatives
  - Often used for demographic information and qualitative questions
  - Typically 2 to 6 response alternatives are used – use more if specific and narrow information is needed
  - If in doubt, best to use a broader set of alternatives – if too broad, they can be grouped during analysis into narrower set of alternatives
  - Choosing one or multiple responses from a list
  - For one choice, instruct the respondent to select the one that is most descriptive
  - For two or more choices, instruct the respondents to check all the answers that apply
  - Use sparingly the "check all that apply" alternative – can cause confusion for respondents if alternatives are not nearly equal in their view
- **Rank Items** format
  - Ask respondents to rank all alternatives – Example: "1" for most to "6" for least important if there are 6 alternatives
  - Or, choose the most, second, and third most important alternative from among 4 or more alternatives
  - Rank items are difficult to design and difficult to answer, so use them sparingly
- Are the response choices exhaustive?
  - Does the question identify and list all possible answers?
  - If not, respondents may leave the item blank or make an incorrect selection
  - If you are not sure that the choices are exhaustive, add the alternative "Other (please specify) ____________________
  - A second option is to use the "Do not know" response alternative
  - If a questions may not apply to some respondents, add the "Not applicable" response alternative
- Are the response choices exclusive? Do the options overlap? If so, the respondent will not know which option to choose.

Figure 1.8 illustrates how a survey question can be structured in different ways.

NOTES AND SUPPLEMENTAL INFORMATION

Table 1.3: Examples of Various Likert Response Formats

Type of Scale	Points within Scale Continuum 5	4	3	2	1
Agreement	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Frequency	Always	Often	About half the time	Seldom	Never
Satisfaction	Very satisfied	Satisfied	Neither satisfied nor dissatisfied	Dissatisfied	Very dissatisfied
Effectiveness	Very effective	Effective	Neither effective nor ineffective	Ineffective	Very ineffective
Quality	Very good	Good	Average	Poor	Very poor
Expectancy	Much better than expected	Better than expected	As expected	Worse than expected	Much worse than expected
Extent	To a very great extent	To a great extent	To a moderate extent	To a small extent	To no extent

Figure 1.9: Semantic Differential Rating Scale Adjectives

ACTIVITY:

Active ___ : ___ : ___ : ___ : ___ : ___ : ___ Passive
Fast ___ : ___ : ___ : ___ : ___ : ___ : ___ Slow
Relaxed ___ : ___ : ___ : ___ : ___ : ___ : ___ Tense
Hardworking ___ : ___ : ___ : ___ : ___ : ___ : ___ Lazy
Efficient ___ : ___ : ___ : ___ : ___ : ___ : ___ Inefficient

EVALUATION:

Good ___ : ___ : ___ : ___ : ___ : ___ : ___ Bad
Valuable ___ : ___ : ___ : ___ : ___ : ___ : ___ Worthless
Clean ___ : ___ : ___ : ___ : ___ : ___ : ___ Dirty
Tasty ___ : ___ : ___ : ___ : ___ : ___ : ___ Distasteful
Useful ___ : ___ : ___ : ___ : ___ : ___ : ___ Wasteful
Sincere ___ : ___ : ___ : ___ : ___ : ___ : ___ Insincere

POTENCY:

Strong ___ : ___ : ___ : ___ : ___ : ___ : ___ Weak
Hot ___ : ___ : ___ : ___ : ___ : ___ : ___ Cold
Helpful ___ : ___ : ___ : ___ : ___ : ___ : ___ Harmful
Too Frequent ___ : ___ : ___ : ___ : ___ : ___ : ___ Too Infrequent
Too Long ___ : ___ : ___ : ___ : ___ : ___ : ___ Too Short
Deep ___ : ___ : ___ : ___ : ___ : ___ : ___ Shallow

STEPS AND PROCEDURES

Table 1.3 illustrates Likert-type scales.

Figure 1.9 is an example of adjectives that can be used in a Semantic differential rating scale.

3-5 The special case of the **Rating Scales**: Used primarily to gather data about beliefs, attitudes, and judgments. Provides data that uses values rather than just categories.

- Likert Rating Scale
  - Most widely used rating scale in questionnaires
  - Respondents are asked the amount they agree or disagree with a number of statements
  - A 'true" Likert scale uses a 5-point scale
  - Categories are (1) strongly agree, (2) agree, (3) undecided, (4) disagree, (5) strongly disagree
  - Additional sets of anchors can be used – they are called Likert-type scales
  - Most questionnaires use a Likert or Likert-type scale of 5 points – some use as many as 11 points – increasing the number of points beyond 5 usually does not help because most respondents are unable to make such fine distinctions
- Should a Likert-type scale have a midpoint?
  - A mid-point allows respondents with a truly neutral attitude a way to respond if respondents are ambivalent on a topic, they may be frustrated if forced to express an opinion where there is no neutral midpoint
  - When computing averages for an item, your mean may be at the midpoint, yet your response scale has no such number
  - If an item is "action-oriented" and several respondents choose the neutral midpoint, what does this mean? In such instances, you may choose to force respondents to choose one side or another by eliminating a neutral midpoint
- Semantic Differential Rating Scale
  - Respondents are asked to think about and judge a particular concept or issue using a set of scales anchored at extreme points by words of opposite meaning
  - Typically, the words used are 4 to 6 sets of bipolar adjectives and the number of scale points is between 5 and 7
  - Respondents mark the place on the scale that best describes their perceptions or attitudes
  - The results of semantic differential scales can be used to assess the respondent's overall reaction to different concepts or issues
  - Particularly useful for checking to see if over time there has been a change in attitude and in what direction
- Semantic Differential is a useful alternative to Likert-type scales if:
  - Respondents have responded to several questionnaires using Likert-type scales and begin to treat the scales in a predictable manner
  - Only a few "attitude" items are being used and the topic involves strong opinions and attitudes

3-6 There are 12 standards for effective item writing. Have a team look at each question and use a checklist to apply to each question.

NOTES AND SUPPLEMENTAL INFORMATION

STEPS AND PROCEDURES

Table 14.4: Item Writing Standards

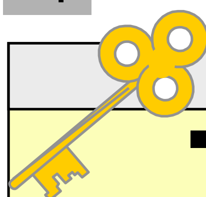
STANDARD	PROBLEM	BETTER
1. Syntax	Equity and economics not with-standing, my compensation is adequate for a satisfactory living situation.	My pay is adequate.
2. Jargon	To what degree are level 1 evaluations reviewed by HRD staff for program enhancement?	How are participant reaction questionnaires from training programs used by human resource development staff?
3. Answerable	To what degree are the sales people adequately informed about the new product line?	Are you informed on the new product line?
4. Loaded Language	Our CEO initiated reengineering program has significantly contributed to cutting our once high costs.	The reengineering program has resulted in cost savings.
5. Negatives	I am not willing to work more overtime.	I am willing to work overtime.
6. Double-Barreled	My office space and equipment is satisfactory.	My office space is satisfactory. My office equipment is satisfactory.
7. Self-Incriminating	I have used work tools for personal business.	Use of work tools for personal business is a problem in the company.
8. Scale Agreement	To what degree does your manager keep agreements? ● Strongly Agree ● Agree ● Undecided ● Disagree ● Strongly Disagree	My manager keeps agreements. ● Strongly Agree ● Agree ● Undecided ● Disagree ● Strongly Disagree
9. Balanced Categories: NOT mutually exclusive:	My hourly pay rate is: ● $13 to $14 ● $14 to $15 ● $15 to $16 and so on	My hourly pay rate is: ● $13.00 to $13.99 ● $14.00 to $14.99 ● $15.00 to $15.99 and so on
NOT exhaustive:	As part of my job, I read: ___ reports ___ memos	As part of my job, I read: ___ company reports ___ industry reports ___ office manuals ___ technical manuals ___ trade journals ___ interoffice memos
NOT balanced:	I like my work: ● Very Strongly Agree ● Strongly Agree ● Agree ● Somewhat Agree	I like my work: ● Strongly Agree ● Agree ● Undecided ● Disagree ● Strongly Disagree

Continued On Page 1-24

STEPS AND PROCEDURES

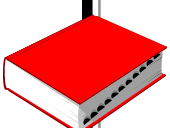
Table 1.4 explains effective item writing standards.

4 Order of Items



Ordering Questions & Front Matter

- The order in which questions are presented can influence response rates.
- A questionnaire MUST begin with an explanation of why the respondent is being asked to participate in the needs analysis study.
- Mailed questionnaires REQUIRE a cover letter.



See pages 1-8 to 1-11 for more guidelines and examples on effective item writing.

4-1 For long questionnaires, break the items up into different sections. Questions concerning a given aspect of the performance analysis should be grouped together.

4-2 Generally, for mailed questionnaires, place the most interesting questions from the viewpoint of the respondents first.

- As the respondents glance over the first few questions, this may motivate them to continue
- Ask about interest in questions during the pre-testing of the questionnaire
- Place less interesting questions such as those about demographics towards the end

5 Put the Questionnaire Together

5-1 Physically layout the questionnaire. Pay attention to the following guidelines:

- Use lots of "white space" – do not clutter the questionnaire
- Seek professional look – consider use of organization logo, distinctive type, color(s)
- Make questionnaire convenient to complete for the user
  - Consider possible places and situations the respondent might be in when completing the questionnaire (e.g., home in an easy chair, cluttered desk at work, cafeteria, in a meeting, etc.)
  - Answer sheet separate from questions can be troublesome
- Keep number of pages small – stapled pages are a problem -- "booklet" look is better (e.g., one 17" x 11" page folded with four face pages better than two stapled pages with same four face pages)
- If items spill over on a second page – repeat the response scale on the second page
- Study other questionnaires as models

HINT: Build a file of questionnaires for reference purposes.

NOTES AND SUPPLEMENTAL INFORMATION

STEPS AND PROCEDURES

Table 1.4: Item Writing Standards (continued)

STANDARD	PROBLEM	BETTER
10. Problem Identification:	Do you feel personally responsible for the success of the organization?	Does your unit know what is expected of it?
11. Tied to Objectives:	If questionnaire objective is about expanding product lines: I tell my friends that this is a good place to be a sales person.	New product lines will increase my customer base.
12. Edit: **Note**: If using computer software, do not rely entirely on a spell-checking program.	The **sails** data **is** accurate.	The **sales** data **are** accurate.

Figure 1.10: Things to Consider in a Cover Letter

- Importance and goals of study – Example: to improve practice
- Impress upon respondents the importance of their opinions and participation
- Explain what will happen with the information
- Explain whether information will be kept anonymous or confidential
- Indicate that if there are questions or concerns, contact ___________ (list name, title, address, telephone number, FAX number, e-mail address, etc.)
- Provide a cutoff date for returning the questionnaire
- Indicate if, where, how, and when the results will be disseminated or published

5-2 Prepare instructions.
- Tell respondents how to mark items on the questionnaire
- Make brief and very clear
- Consider an example if response categories are complex or unusual
- If response categories change – write instructions for each section

5-3 Prepare scales.
- When feasible, best to use one scale throughout a questionnaire
- When not feasible to use one scale – make them distinct
- Consistently display the scale items
  - Make marking of items clear and easy
  - Use check of boxes or have items circled, etc.

Review pages 1-11 and 1-13 for information on questionnaire layout.

5-4 Prepare check off items – demographic items, etc.
- Specify if respondent is to “check only one item” or “check all that apply”
- Provide a place for the check mark – boxes, circles, lines, etc.

5-5 Prepare open-ended questions.
- Ask for specific comments
- Indicate space to be used – and indicate expected length of the response
- Strategically place open-ended questions at end of:
  - Major sections or questions that may need elaboration
  - Questionnaire for “other comments”

Figure 1.10 recommends things to consider in a cover letter.

5-6 Indicate what the respondent is to do with the questionnaire.
- Place these instructions in the cover letter
- Repeat instructions at the end of the questionnaire

5-7 Prepare the cover letter.
- Can be separate or first page of questionnaire – may help to use official stationary
- Should be signed by senior management (also, consider champions)

5-8 As with individual items, carefully edit other parts of the questionnaire.

6 Select Respondents

See section 2 on populations and samples for details on selecting a sample.

6-1 Define the target population. Generally, this is ALL individuals who have knowledge and views relevant to the content of the questionnaire. Reviewing the statement of purpose will help in defining the population.

6-2 If the population is small (i.e., 500 or less), consider conducting a census – giving the questionnaire to everyone in the population.

6-3 For larger populations, use a sample of the population.

NOTES AND SUPPLEMENTAL INFORMATION

STEPS AND PROCEDURES

Figure 1.11: Questionnaire Pre-Test Instructions

Explain that pre-test purposes are:

- To evaluate the questionnaire
- Check out clarity of introduction and instructions
- See if wording of items, response alternatives, and scales are clear
- Check out comprehensiveness of item alternatives

Ask the pre-test respondents to:

- Draw circles around anything they find confusing, incomplete, or not applicable
- Complete the survey at the speed they would normally use to complete a questionnaire

Keep a record of:

- The amount of time it takes each respondent to complete the questionnaire
- Either note the time or have the respondent write down the time they completed the questionnaire

STEPS AND PROCEDURES

7 Pre-Test and Revise the Questionnaire

Pre-Testing Questionnaires

- Pre-testing a questionnaire is a multistage and cumulative process.
- What is clear to a questionnaire developer is often NOT clear to its intended audience.
- There are two purposes to pre-testing.
  - To assess the content of the questionnaire.
  - To get an estimate of the amount of time it takes respondents to complete the questionnaire.

Performance analysis questionnaires MUST BE pre-tested. Tryout items and the process.

7-1 For questionnaire pre-testing, purposely choose people who are the same as those who will receive the completed questionnaire. The circumstances under which they respond to the pre-test should be similar as well.

- 8 to 12 people is a reasonable size for a pre-test group
- Use a diverse sample for pre-testing – people who reflect the characteristics of the entire group
- If incremental changes are made in the questionnaire, use the same sample of people for each subsequent phase of pre-testing

7-2 Explain to the respondents the purposes of the pre-test and give them instructions you want them to follow.

Figure 1.11 lists instructions to give to respondents for conducting a questionnaire pre-test.

7-3 Conduct a debriefing. Focus on feedback from the respondents concerning the following information:

- Positive and negative feelings about the questionnaire
- Instructions or items that were unclear, difficult to understand or follow, or offensive
- Suggestions for topics, items, or alternatives to add to the questionnaire

7-4 Make revisions to the questionnaire.

- If the time taken by the pre-test group is too long, shorten the questionnaire
- Revise instructions, items, and layout based upon pre-test feedback

7-5 If substantial changes are made to the questionnaire, repeat the pre-testing process.

NOTES AND SUPPLEMENTAL INFORMATION

STEPS AND PROCEDURES

Table 1.12: Mailing Suggestions to Maximize Returns

- Cover letter should have date and handwritten signature.
- Use postage paid self-addressed envelope. Respondents are not likely to furnish their own envelope or postage.
- Send in a business-size envelope.
  - Individually type or laser-print addresses
  - DO NOT use prepared mail labels if seeking a 'professional' look
- Use first-class postage -- advantages:
  - Gets through postal system faster
  - It's forwarded or returned if necessary
  - It looks important
- Disadvantages of bulk-rate postage:
  - Slower and sometimes gets lost by postal service
  - Gets discarded if problems with delivery
  - Many people throw out bulk mail as "junk mail"

STEPS AND PROCEDURES

Converting Open-Ended Questions

- For open-ended questions, the pre-test may yield a definitive set of responses that will allow for converting the item from an open- to a closed-ended question.
- If this is done, have the item undergo another pre-testing.

8 Administer the Questionnaire

8-1 Reproduce the questionnaire -- estimate the number needed and multiply by 1.5 or 2.0.

- If the questionnaire is mailed, consider the following:
  - Use a professional look -- do not want to look like “junk mail”
  - Consider printing a message on the envelope -- “Your opinions will help………”
  - Prepare a follow-up letter -- you will have non-respondents -- let them know via the letter that you know they are busy, but emphasize the importance of returning the questionnaire
- Reproduce appropriate accompanying materials
  - Cover letter
  - Mail out envelopes
  - Return envelopes
- Assemble materials for dissemination

Extra copies of questionnaires are needed for follow-up mailing, coding, for reports, etc.

8-2 Disseminate the questionnaire.

- If the questionnaire is administered to respondents gathered in a group setting, consider the following:
  - Provide the questionnaire administrator with notes to use in verbally explaining the purposes of the questionnaire and to answer questions
  - If responses are to be truly anonymous or confidential, have the questionnaire administrator explain how anonymity or confidentiality will be assured
  - Make provisions to provide participants with pencils or pens and physical space to comfortably complete the questionnaire
- If the questionnaire is administered to the respondents by mail, consult with U.S. Post Office or company mail services concerning:
  - Class/type mail service to use
  - Application of outgoing postage and postage on return envelopes
  - Bundling the out-going envelopes for bulk mailing
- Assemble personnel and prepare mailing

Table 1.12 offers Mailing Suggestions

Figure 1.13: Follow-Up Mailings

ONE WEEK **Reminder postcard**

- Thank you to those who responded
- A brief message to non-respondents indicating the importance of the questionnaire and a request that it be completed and returned immediately

THREE WEEKS **Replacement questionnaire to non-respondents**

- New personalized cover letter with handwritten signature and a questionnaire with return envelope
- Make the tone insistent regarding the importance of the questionnaire and a personal appeal letting them know you are waiting on their response.

SEVEN WEEKS **Certified mailing to non-respondents**:

- If more returns are necessary to get a representative sample send another personalized letter with handwritten signature and a questionnaire with return envelope
- Make the tone demanding and let them know that you expect them to return the questionnaire

Figure 1.13 lists guidelines for follow-up mailings.

8-3 Monitor returns of questionnaires and code for data processing.

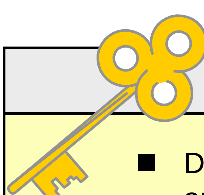
- Record number returned on a daily basis
- Open returned envelopes and edit and code returned questionnaires
- Check early sample of returned questionnaires for problems
  - Unanswered questions
  - Pages missing or stuck together
  - Instructions incorrectly followed

8-4 Carry out follow-up mailings.

8-5 If, after initial and follow-up mailings, the return rate is small, consider the following:

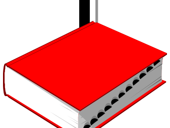
- Contact via telephone or in person random sample of non-respondents
- Ask them why they did not return the questionnaire
- If agreeable, ask them to respond to the questions verbally
- Record responses

9 Analyze the Data



Data Analysis Guidelines

- Do not confuse significance with importance -- are differences and trends in data different enough to be important?
- Build good tables, graphs, and charts to help in “visualizing” the data
- For most cases (90% of the time), analyze your data using:
  - Measures of central tendency: means, medians, and modes
  - Measures of dispersion: ranges, standard deviations, and normal or skewed distribution
  - Cross tabulations
- If data “hit you between the eyes” -- you probably have findings of practical importance



See Parts 3, 4, and 5 for details on data analysis.

9-1 Count and total answers to:

- Demographic items -- calculate as percentages
- Closed-form items -- calculate as percentages

9-2 Analyze the content of the open-form items.

- Categorize responses
- Count and calculate percentages as appropriate
- Note responses which exemplify a cluster of like responses

NOTES AND SUPPLEMENTAL INFORMATION

STEPS AND PROCEDURES

Table 1.5: General Guidelines for Graphically Displaying Questionnaire Data

Type of Data	Definition & Examples	Graphical Representation
Nominal	Distinguishable names, labels, or categories: ■ Gender: male or female ■ Employment Status: full or part-time ■ Shift worked: first, second, or third	**Bar Graph** -- number in each category **Pie Chart** -- percentage in each category
Ordinal	Distinguishable and rank ordered: ■ Job Ranking: paint manager, supervisor, or shift leader ■ Shirt size: small, medium, large or extra large ■ Age Categories: under 18, 18-25, 26-30, 31-35, 36-40, 40 and over	**Frequency table** -- number or percentage in each category **Histogram or polygon** -- distribution of scores or percentages
Equal Interval	Classify along a numerical continuum with equal distance among the values: ■ Scores on an objective test ■ Opinions/beliefs measures on a Likert or Likert-type scale (SA A U D SD) ■ Hours worked or money earned	**Tables** -- with frequencies and percentages **Line or Area Charts** -- frequency on vertical axis and scores or percentages on horizontal axis.

Table 1.5 illustrates general guidelines for graphically displaying questionnaire data.

9-3 Decide on any comparisons to make.
- Demographic items to response items
- Combinations of response items
- Prior survey information
  - Responses at different time periods
  - Compare responses from mail to those of sample of "non-respondents" that were interviewed and note similarities or differences

9-4 Draw tables, graphs, and charts to "visualize" data.

9-5 Carry out the following descriptive or inferential statistical analyses as deemed meaningful.
- Cross tabulations between data in comparison tables
- Correlation between two or more variables
- Differences between two or more groups
- Conditions which lead to certain attitudes, actions, or conditions
- Combining of responses to a number of questions to form indices

  NOTE: consult statistical expert as needed

9-6 Sketch out findings.
- Look for gaps and missing information
- What needs, resources, and barriers emerge
- Make certain evidence supports findings

9-7 Draw conclusions based on findings and evidence collected.

10 Report Findings

10-1 Identify audience for report.
- Who will read the report?
- What information will the audience need/use?
- Under whose name will the report be published?

10-2 Write introduction.
- Context -- background for the study
- Purpose -- why the study is important
- Questionnaire objectives -- be specific

10-3 Describe methods used.
- Respondents
  - Describe population
  - Describe sample -- how selected and how many responded
- The questionnaire
  - How was it developed?
  - Explain tryouts and pre-testing
  - Describe major clusters of questions
  - For short questionnaires -- append a copy
  - For long questionnaires -- append examples of items

NOTES AND SUPPLEMENTAL INFORMATION

STEPS AND PROCEDURES

Figure 1.14: Outline for a Questionnaire Final Report

Check when completed

☐ **Title Page** Date Completed ____________	☐ Title for study ☐ Sponsoring group (champions) ☐ Name, address, and phone number of contact person(s) ☐ Date of publication
☐ **Table of Contents** Date Completed ____________	☐ Outline of main topic headings ☐ Outline of subtopic headings
☐ **List of Illustrations** Date Completed ____________	☐ Figures ☐ Tables ☐ Charts
☐ **Introduction Or Forward** Date Completed ____________	☐ Why the study was conducted ☐ Problem statement ☐ Group(s) involved ☐ Sponsors (champions)
☐ **Background** Date Completed ____________	☐ Purpose(s) and goal(s) ☐ Description of population and sample ☐ Sampling method(s) used ☐ Responses and non responses ☐ Time frame(s) – including dates
☐ **Findings** Date Completed ____________	☐ Verbal explanations of what is found in figures, tables, charts, etc. ☐ Verbal explanations of anecdotal data ☐ Significant comparisons of data ☐ Overall picture(s) revealed by the data ☐ Representative quotations from open-ended items ☐ Conclusions and recommendations
☐ **Summary** Date Completed ____________	☐ Concise abstract of important findings and recommendations
☐ **Appendixes** Date Completed ____________	☐ Copy of questionnaire ☐ Bibliography ☐ Glossary of terms ☐ Letter of endorsement or authorization ☐ Tabular material – data summaries

STEPS AND PROCEDURES

Figure 1.14 is an example of a comprehensive outline for reporting questionnaire findings.

- Procedures
  - When and how distributed and returned
  - Explain issues of informed consent and/or anonymity or confidentiality
  - Data analysis techniques used
- Weaknesses -- be frank about:
  - Low return rates
  - Bias in returns
  - Items that may have confused respondents

10-4 Present results.

- Use tables, graphs, and figures to display findings
- Explain major/important findings in straightforward language
- Reference table, graphs, and figures in text

10-5 Write discussion and conclusions

- Summarize and highlight major results
- Indicate implications of findings -- what should be done
- Tie information together and reach conclusions
  - Stay within the parameters of the data
  - Consider referring to prior studies or other data sources as appropriate

10-6 Prepare front and back matter

- Title page - sponsoring group - name and address of contact person
- List of tables, graphs, and figures
- Appendix considerations:
  - Copy of questionnaire or example items
  - Bibliography
  - Letter of endorsement or authorization
  - Glossary of terms
  - Tabular material not summarized

10-7 Prepare Executive Summary.

- General summary for quick and easy reading
- Highlights findings and recommendations

USCG

WORKSHOP

U
R
V
E
Y

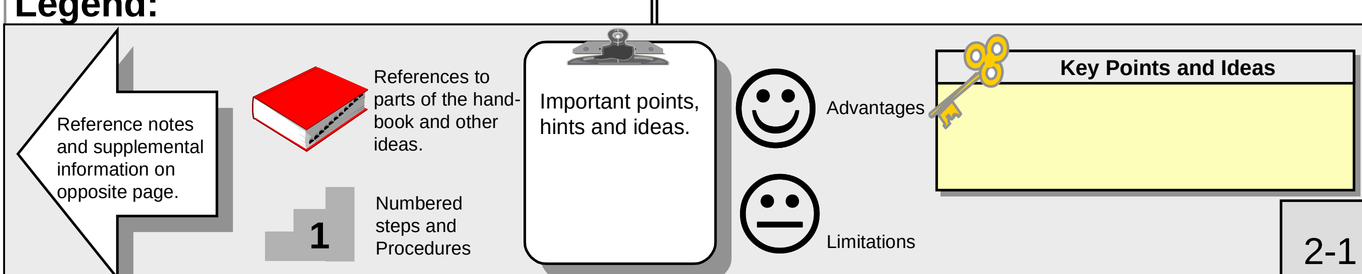
Section 2: Populations and Samples

CONTENTS

Overview

Guidelines

Legend:



NOTES AND SUPPLEMENTAL INFORMATION

OVERVIEW

Table 2.1: Commonly Used Probability Sampling Methods

Method/Description	Benefits	Concerns
Simple Random Sampling Every unit or case in a population has an equal chance of being selected.	/ Relatively easy to do. / Printed tables and computer programs are readily available to generate random selections. / For relatively small populations, one can literally pull numbers or names "out of a hat."	/ If there are subgroups in a population that are of particular interest, they may not be included in appropriate proportions in the sample. / If the population is large and organized in lists, it is labor intensive to assign numbers to the list and identify randomly chosen cases.
Systematic Sampling Every *n*th unit or case on a list is selected. *N*th for example can be every 5th, 10th, or 25th case. It is determined by dividing the size of the popula-tion by the desired sample size.	/ Very convenient to use existing lists of people, objects, events, etc. / Similar to random selection if the starting point in a list is randomly chosen.	/ If there is a recurring pattern within a list, the method may yield bias results. For example, data arranged by month or lists of names that are grouped with the group leaders name always first are problematic.
Stratified Random Sampling The population under study is grouped by meaningful characteristics (i.e., strata). Every unit or case in each strata has an equal chance of being selected.	/ Easy to conduct analysis of pertinent subgroups such as gender, age, experience, geographic location, etc. / The sample is more likely to reflect the population than with simple random sampling.	/ Must calculate population and sample sizes for each sub-group. / Can be labor intensive and difficult if several subgroups are utilized -- particularly if individuals or cases in the subgroups are not readily identifiable.
Cluster Sampling Natural groups or clusters are sampled with members of each selected cluster being sub - sampled.	/ Convenient to use existing intact units. / Clustering concentrates data collection within fewer and smaller settings reducing data collection time and costs. / It is useful when it is either impossible or impractical to compile an exhaustive list of the units that make up the target population.	/ Cluster sampling is efficient, but it yields a less accurate sample. / General assumption is that all of the clusters are alike. If they are quite different, the sample will be biased. / It is best to maximize the number of sampled clusters while decreasing the number of units within each cluster.

The decisions a performance analyst makes regarding the people, objects, groups, or organizational entities to study are important. Choosing what to study and whether or not to include all or only part of the individual elements that make up the study group falls into the realms of populations and samples.

Introduction

Performance analysis studies often begin with a general idea about the people, things, or events to be studied. These are called the ***unit of analysis***. Through prior projects and experience as well as from information in the professional literature, the analyst develops a conceptual definition of the unit of analysis to be studied. The analyst specifies whether the unit of analysis is made up of individual units or groupings of units. For instance, individual production workers or production teams in Factory XYZ.

Populations and Samples

An entire group of people, things, or events being studied is a ***population***. A population is defined by identifying the unit of analysis. The definition requires specifying the common characteristics of the individual objects or elements which make up the collection of people, things, or events. This is done in such a way that the analyst can discriminate between objects or elements which belong to and those which do not belong to the collection. All of the employees of a company, all of the company's meeting rooms, and all of the company's pay days in a given year are examples of populations.

A single person, thing, grouping, or event is called an ***element*** of a population. Pat, a company employee, the safety training room, and the last pay day of the first calendar quarter are examples of elements of populations.

A selection of one or more elements from a population is called a ***sample***. A sample can range in size from one element in a population to all of the elements in a population, but one. Simply, a sample is a ***subset*** of a population.

Why Sample?

Often limited resources or time make it unpractical to study a population. Sometimes the population is very large or all of the elements are not available for study. In such instances, a sample of a population is studied.

An example of a large population would be the 10,500 wholesale customers who contacted a company by phone last calendar year with some no longer in business. A not yet available population would be 60 new automated telephones to be installed over the next six month period -- 10 per month.

Table 2.1 describes the benefits and concerns associated with four popular probability sampling methods.

Two Sampling Strategies

Sampling strategies can be divided into two categories -- ***probability*** and ***non-probability.*** Probability sampling involves ***random selection***. Random means selection of an element or grouping of elements independent of the selection of any other unit. Commonly used probability sampling

NOTES AND SUPPLEMENTAL INFORMATION

OVERVIEW

Table 2.2: Commonly Used Non-Probability Sampling Methods

Method/Description	Benefits	Concerns
Convenience Sampling Using a group of units or cases that are readily available.	/ A practical method that relies on readily available people, objects, or events. / Inexpensive and time efficient.	/ The sample is opportunistic and the units or cases chosen are unlikely to be representative of the target population.
Purposive Sampling The analyst relies on his/her judgments to select units or cases that represent or are typical of the population.	/ Helpful in exploratory or developmental work. Makes certain the widest variety of subjects are represented. Useful in comparing and contrasting extremes in a population.	/ Requires full knowledge on the part of analyst concerning the relevant factors that vary in a population. Easy for the bias of the analyst to negatively impact the representativeness of the sample or for the analyst to be incorrect in identifying relevant factors which vary in the population.
Quota Sampling The population is divided into subgroups and the sample is based upon the subgroups needed to represent their proportions in the population. Samples from the subgroups are chosen using either convenience or purposive sampling.	/ Practical if reliable data are available to describe the proportions of units or cases in the various subgroups.	/ Records of the proportions of the subgroups must be accurate and up-to-date. / Sampling bias is a concern, but the method will yield a more representative sample than convenience or purposive sampling.

OVERVIEW

methods include ***simple random sampling***, ***systematic sampling***, ***stratified random sampling***, and ***cluster sampling***. Each method has it benefits and limitations. The essence of probability sampling is to choose a sample that fully represents the population under study. The performance analyst is seeking to generalize from the sample to population as a whole.

Table 2.2 describes the benefits and concerns associated with three popular non-probability sampling methods.

Non- probability sampling involves selecting elements from a population by means other than randomly. For most non-probability sampling methods the analyst makes judgments about the purposes of the study and the nature of the population and then purposefully chooses certain elements to make up the sample. Commonly used non-probability sampling methods include ***convenience sampling***, ***purposive sampling***, and ***quota sampling***.

Population or Sample?

Ideally, it is desirable to collect data from a population. In so doing, there is no need to make inferences from a sample to the population about what is found. In short, “what you get is what you got” in terms of data. No guesswork is needed about how representative the data are of the population. Population studies are realistic if the size of the population is relatively small, all elements of the population are accessible, and adequate time/resources are available.

For most performance analysis studies, the analyst can make reasonable choices about the population to be studied and the sampling strategy to use. The guidelines which follow will help in making good choices. If a study is large, involves many resources, and the findings will be consequential, the analyst should consider seeking technical assistance.

Often in needs analysis studies, the population is large, difficult to fully access, and time and resources are limited. Such cases require sampling. If the needs analyst is interested in making certain that data findings from a sample can be generalized to the population as a whole, the sampling goal will be to choose a sampling strategy that will yield a sample which closely approximates the population in characteristics important to the study. Important characteristics may include demographics, work related factors, socio-economic factors, or time and location.

When sampling, several factors come into play in choosing a strategy. The stage of the study, how the data will be used, time and resources, and method(s) of data collection come into play. The analyst must choose a degree of confidence -- that is certainty -- that the sample represents the population. Finally, the precision desired for planned sample statistics -- that is, the margin of error -- must be considered. All of these factors are explained in detail in the following guidelines.

Further Readings

Babbie, E. (1995). The Practice of social research (7th edition). Belmont, CA: Wadsworth.

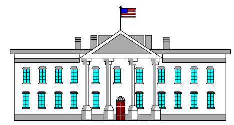
Edwards, J. E., Thomas, M. D., Rosenfeld, P., & Booth-Kewley, S. (1997). How to conduct organizational surveys: A step-by-step guide. Thousand Oaks, CA: Sage.

Mertens, D. M. (1998). Research methods in education and psychology: Integrating diversity with quantitative & qualitative approaches. Thousand Oaks, CA: Sage.

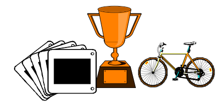
NOTES AND SUPPLEMENTAL INFORMATION

Figure 2.1: Units of Analysis (objects or cases being analyzed)

GROUPS
- Safety Teams
- Second Shift Workers
- Regional Sales Forces
- Quality Improvement Teams
- Production Supervisors
- Secretarial Pools



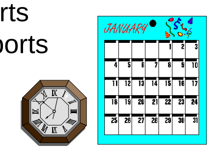
ORGANIZATIONAL UNITS
- Retail Stores
- Manufacturing Plants
- Regional Operations
- Western Sales Divisions
- Maintenance Departments
- Supply Co-ops



ARTIFACTS
- Memos/Letters
- Finished Products
- Financial Statements
- Machines
- Meeting Rooms
- Annual Reports
- Accident Reports

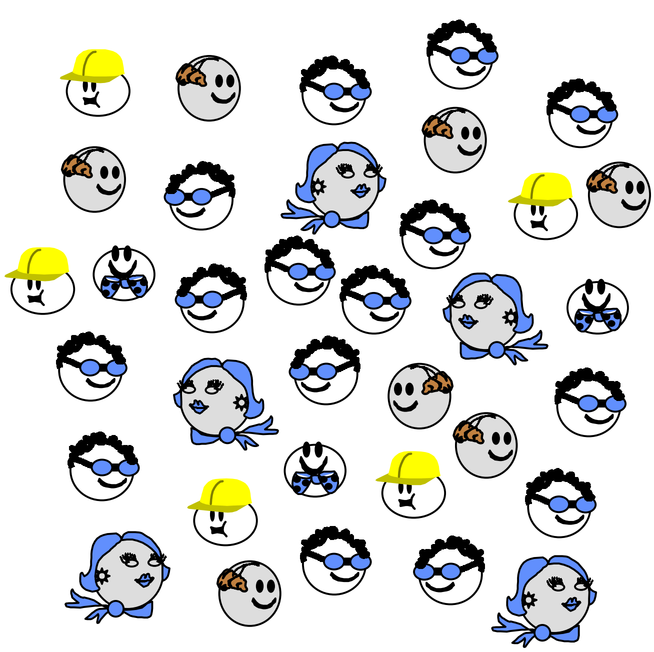
PEOPLE
- Hourly Workers
- Salary Workers
- Trainees
- Sales Staff
- Customers
- Clients
- Interns

EXAMPLES



OCCURENCES
- Holidays
- Pay Days
- Calendar Quarters
- Promotional Sales
- Staff Meetings

Figure 2.2: Population of Workers From a Small Company



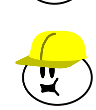
A Population of 35 Workers

 3 Managers

 5 Sales People

 7 Office Support Workers

 15 Production Workers

 5 Delivery Drivers

GUIDELINES

Introduction

In designing a performance analysis study, the first question to answer is: "What is it that is going to be studied?" The persons, things, or events that are the focus of a performance analysis study are called the "units of analysis." The units of analysis are the objects or cases of the study. These may be people, groups, occurrences, organizations, or artifacts.

Figure 2.1 shows different types of units of analysis used in Performance analysis studies.

What Should Be the Unit of Analysis?

The purpose of the study dictates whom or what is to be described, analyzed, and compared. It is important to carefully determine the units about which you wish to draw conclusions and then make sure that the data you collect pertain to those units. For example, if you are interested in the safety records of individual workers and the data that are available combine accidents for all individuals by department, then conclusions about individuals must be drawn very tentatively.

Individual Versus Aggregate Data

It is important to clarify and keep in mind during a performance analysis study whether the unit of analysis is individuals or groups (aggregates). When information about individuals is combined to describe the characteristics of a group to which they belong this can be either individual or aggregate data, depending on how the data are used. If collective information about a group is used to describe the characteristics of individuals within the group, the unit of analysis is the individual. However, if aggregate information about a group is used to compare different groups, then the unit of analysis is the group.

The **ecological fallacy** deals with drawing conclusions about individuals based solely on the observation of groups. Observed group patterns may be genuine, however, this does NOT warrant making assumptions about the causes of the patterns. In short, assumptions cannot be made about individuals making up the groups.

The reason that it is important to accurately identify the unit of analysis is that confusion over the unit may result in false conclusions about performance analysis data. Generally, conclusions about performance analysis data can only be made about the unit under study. For example, one of the most common fallacies in mixing units of analysis is when characteristics of groups are used to make inferences about the individual behaviors of the people within those groups. This is called the "ecological fallacy." For example, if accident rates among workers are higher in departments with a high concentration of older workers, you would be committing an ecological fallacy if you concluded based upon the group data that older workers are more likely to have accidents.

Figure 2.2 is an example of a population of workers.

Populations and Samples

Once the unit of analysis has been determined, the next step in conducting a performance analysis study is to ascertain the population or sample from which to gather data. A population is the total membership of a defined group of people or artifacts.

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

Figure 2.3: Populations and Samples Overview

What is a Population?

	Definition	Examples
What is a Population?	■ TOTAL membership of a defined grouping -People -Groups -Organizations -Artifacts -Occurrences ■ Ability to discriminate -Who/What is in -Who/What is out	■ Sales force for XYZ manufacturing ■ Quality enhancement teams for Widget Manufacturing ■ Regional offices of Ring Telephone Co. ■ Rooms used for training for Worldwide Air ■ Departmental safety records for Fast Transportation Co. ■ Paid holiday for Good Food hourly workers

	Definition	Examples
What is a Sample?	■ Subgroup or part of a population ▪ **Two-Types** 1. <u>PROBABILITY</u> ✓ Random selection ✓ Equal chance of being selected 2. <u>NON-PROBABILITY</u> ✓ Non-random selection ✓ Purposeful choice	■ 100 sales people ✓ Any part of total ✓ Groupings of 1, 2, 3, ….., 99 people 1. <u>PROBABILITY</u> ✓ 100 sales people ✓ Draw 35 names "out of a hat" 2. <u>NON-PROBABILITY</u> ✓ 100 sales people ✓ Select 25 people from Midwest Plant

GUIDELINES

Figure 2.3 provides summary information and examples of populations and samples.

The summarized results of studying a population are called ***parameters***.

A sample is a subset of cases selected from a population. The summarized results of studying a sample are called ***statistics***.

Ideally, performance analysis data are gathered from entire populations. This is the preferred method if the population is relatively small. However, gathering data from an entire population is oftentimes unrealistic because of practical considerations such as limited time and financial resources. Therefore, a sample that accurately represents the variability in a given population is often preferred over gathering data from the entire population. Carefully selected samples yield very accurate results, thus eliminating the need to administer a performance analysis study to an entire population.

There are several different sampling designs from which to choose. Factors such as population size, the time one has to administer the study, resource limitations, and accessibility to the population will impact the decision of which sampling design to select.

Sampling Designs

Ideally, samples should represent the population of interest. This means that the sample should provide a close approximation of certain characteristics of the population. Examples of characteristics include demographic factors such as gender and race or work-related factors such as training and work experience. Sampling designs are generally divided into two broad classes: probability and non-probability.

DEFINITION OF PROBABILITY SAMPLING: sampling based on a process of random selection, which gives each case in the population an equal chance of being included in the sample.

Probability sampling designs always involve the process of 'random selection' at some stage. The primary benefit of the random selection of cases from a population is that you are more likely to have a representative sample. This means that you can assume that the findings from the sample data can be generalized to the entire population.

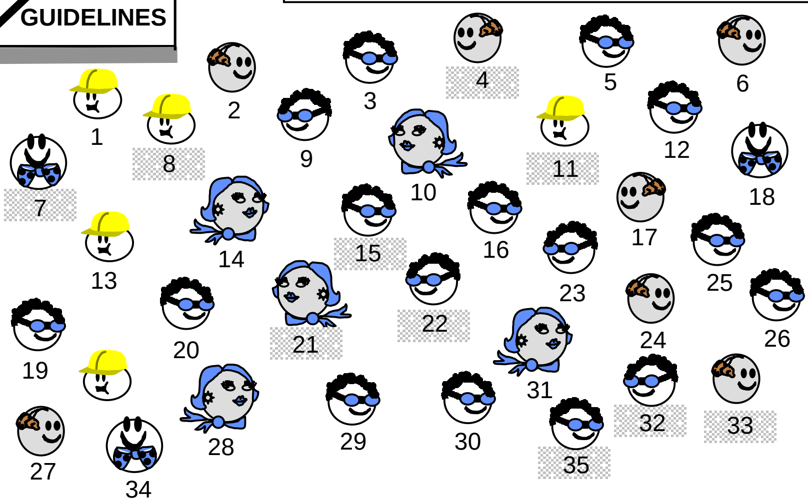
DEFINITION OF NON-PROBABILITY SAMPLING: process of selecting cases other than through random selection.

Non-probability samples have two basic weaknesses: 1) investigator bias in the selection of cases is not controlled and 2) generalizing research findings to the entire population must be done with care. Despite these weaknesses, non-probability sampling is oftentimes either more practical or the only viable means of selecting a sample. .

GUIDELINES

NOTES AND SUPPLEMENTAL INFORMATION

Figure 2.4: Simple Random Sample



Begin

Table of Random Numbers

1290 8766	8874 2686	5847 9338
4457 2304	7209 7333	6942 3208
6574 3490	3352 8948	3329 0732
2395 8422	9402 0398	8584 7294
0203 9485	5940 3094	1284 8739
2019 3847	4633 8211	0193 8474
7382 9215	0382 9746	5929 0121
6388 9215	7238 3172	7141 9131
3377 6797	4451 6767	5288 7431
0927 4236	4633 9707	5353 9773
3569 7935	2576 5458	6438 5481
3738 5687	2328 6166	4672 7230

SAMPLE = 10

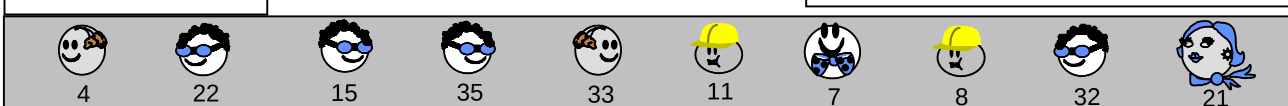


Figure 2.5: Systematic Sampling (alphabetical list of names of 35 employees)

Abe	13 **Dave**	Jake
Abby	Dale	Jane
3 **Arin**	Doug	Janet
Art	Earl	28 **Karen**
Alice	Fran	Kyle
Bart	18 **Frank**	Lee
Barb	Gail	Marc
8 **Bart**	Gus	Nancy
Bruce	Harry	33 **Randy**
Cal	Helen	Wanda
Carl	23 **Irene**	Will
Chris	Irma	

Population Size = 35
Sample Size = 7

Draw one number out of a hat that contains the numbers from 1 to 5. Let's say the number selected is "3".

Count down the list of names to the third name (Arin).

Given the sampling interval is "5", count down every 5th person to get the desired 7 person sample.

$$\textbf{Sampling Interval} = \frac{\textbf{Population Size}}{\textbf{Sample Size}} = \frac{35}{7} = \mathbf{5}$$

$$\textbf{Sampling Ratio} = \frac{\textbf{Sample Size}}{\textbf{Population Size}} = \frac{7}{35} = \mathbf{1:5}\ (20\%)$$

GUIDELINES

Probability Sample Designs

Figure 2.4 illustrates simple random sampling.

Choosing numbers from a table of random numbers is simple. Begin by closing your eyes and sticking a pencil point into the table. Where it lands is where you begin. If you are selecting 10 of 35 subjects, use the last two digits of the beginning number. Proceed to the next number using its last two digits.

Continue until you have 10 random numbers. If a two digit number exceeds the number 35, or a number is repeated, ignore it and move to the next number in the table.

Figure 2.5 illustrates systematic sampling.

1. SIMPLE RANDOM SAMPLING

Simple random sampling adheres to the general principle that every possible combination of cases has an equal chance of being selected as the sample. For example, in a population of four cases, identified as A, B, C and D, there are six possible samples of two cases: (A,B), (A,C), (A,D), (B,C), (B,D), and (C,D). In a simple random sample of two cases, each of these pairs of cases would have the same chance of selection. In order to guarantee this type of random selection, two requirements are necessary:

-- a complete list of the population, and

-- a process for randomly selecting individual cases to be included in the sample.

In order to select a random sample, the needs analyst cannot simply choose cases haphazardly. Preferred methods of random selection typically require the use of computer programs or prepared tables that can be used mechanically to generate random numbers.

2. SYSTEMATIC SAMPLING

Systematic sampling consists of selecting every *n*th (e.g., 5th, 10th, or 20th) case from a complete list or file of the population, starting with a randomly chosen case from the “n” cases on the list (e.g., a randomly chosen case from the first 5, 10, or 20 cases). This procedure has two requirements: a sampling interval (*n*) and a random starting point. The sampling interval is simply the ratio of the number of cases in the population to the desired sample size. A random start refers to the process of randomly generating (generally with a statistical package) an initial case as the starting point between 1 and *n*.

For example, let's assume that you want to select a sample of 100 cases from a total population of 2500. Dividing 2500 by 100, you obtain a sampling interval of 25. You would then select at random a number between 1 and 25, and starting with that number, select every consecutive 25th case. For example, let's assume the random number turns out to be 19. Your sample would then consist of cases numbered 19, 44, 69, ... , 2494.

Systematic sampling is oftentimes preferred over simple random sampling when the population list is quite long and the desired sample is large. Working systematically with a list is more efficient than drawing a simple random sample from a list.

Figure 2.6: Proportionate Stratified Random Sample

Directions

1. Break the population into two or more natural strata.
   Example: - personnel by product lines
   - manual and automatic widgets

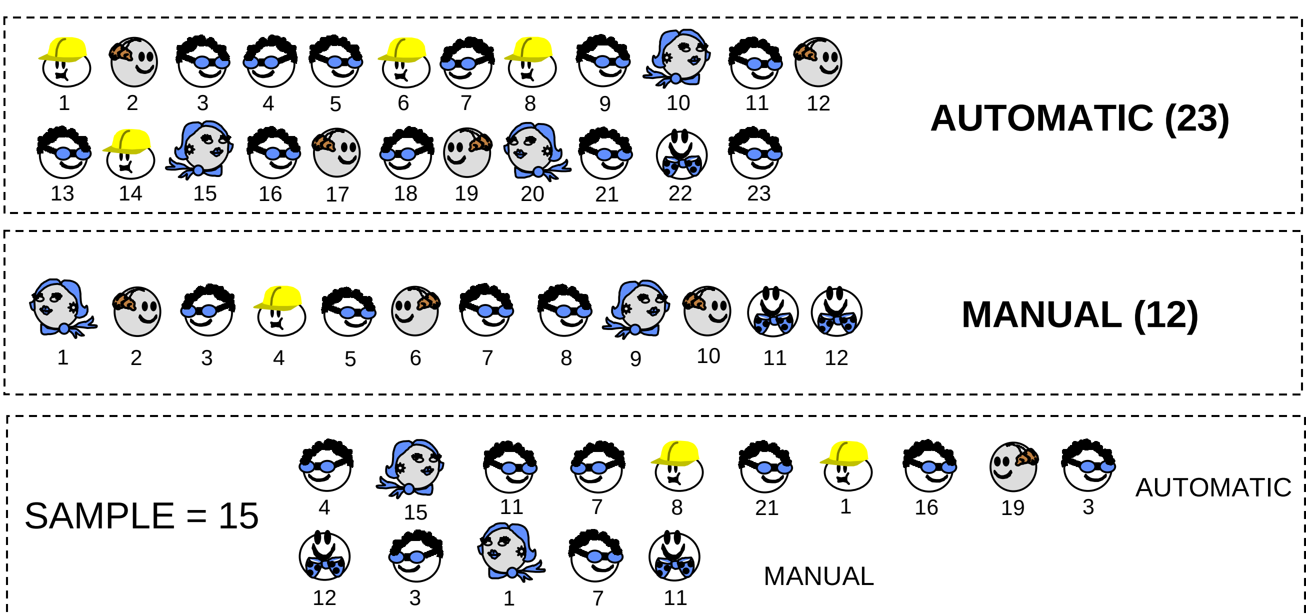
2. Calculate the percentage of the population in each stratum.
   Example: - automatic widget division = 23 of 35 = 65.7%
   - manual Widget division = 12 of 35 = 34.3%

3. Select a sample size.
   Example: - the decision is to select a sample of 15 individuals

4. Calculate the sample size for each strata by applying the percentage for each strata in the population to the sample size.
   Example: - automatic widget division 65.7% of 15 = 9.9 or 10
   - manual widget division 34.3% of 15 = 5.1 or 5

5. Randomly select the required sample from each strata.
   Example: - 10 of the 23 persons in the automatic widget division
   - 5 of the 12 persons in the manual widget division

NOTE: Use a table of random numbers or “pull names out of a hat.”



GUIDELINES

Although systematic sampling is cost and time efficient, the drawback of this sampling design is the possibility of selecting a biased sample. This happens when a population list has a periodic or cyclical pattern that corresponds to the sampling interval. For example, in an organization, if an employee list is arranged such that every *n*th case was a supervisor, your final sample would be comprised solely of supervisors; other employees would not be represented in the sample. Fortunately, this type of cyclical pattern within population lists is relatively uncommon, but it is certainly something to watch for.

3. STRATIFIED RANDOM SAMPLING

In stratified random sampling, the population is first subdivided into two or more mutually exclusive segments, called "strata". Strata are based on categories of one or a combination of relevant variables. Simple random samples are then drawn from each stratum and combined to form the complete, stratified sample. It is used when the analyst wants to ensure adequate representation of subgroups that have particular importance to the issue being analyzed. Some common stratification variables in needs analysis studies are gender, age, race, organizational level, experience level, and operational units.

For example, let's assume that we have a population of eight cases, identified as A through H. Now, let's assume that we are interested in selecting a sample size of four that equally represents cases based on the variable of gender. The cases would first be broken down into two strata: male and female. Now we have two strata of cases that are divided in the following manner: cases A, B, C, and D represent females and cases E, F, G and H represent males. After dividing the cases into two strata, we would then randomly select two cases out of each stratum. The two randomly selected cases from each stratum would then be added together to produce the final stratified random sample. This process would not produce a simple random sample since there are several combinations of cases that are not acceptable, for example, the combinations of all females (ABCD) or all males (EFGH). There are, however, thirty-six possible stratified random samples of four cases with equal proportions of males and females in the samples.

Figure 2.6 illustrates proportionate stratified random sampling.

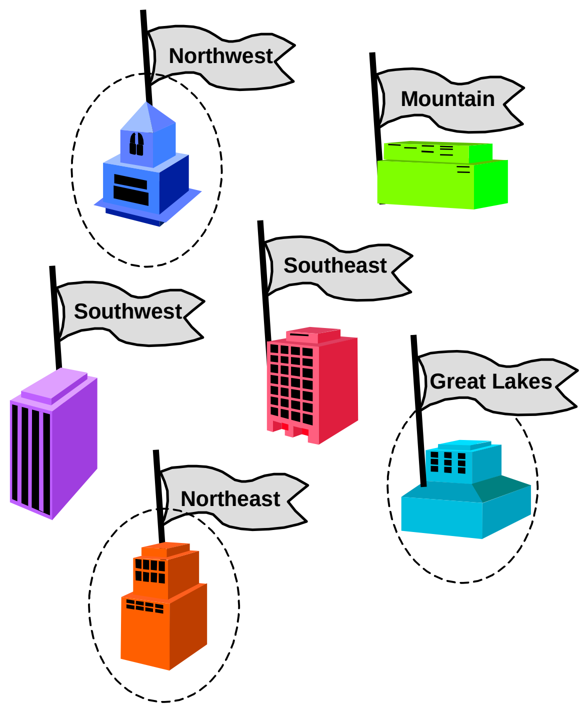
There are two approaches to stratified sampling. The first is ***proportionate stratified random sampling*** where the percentage of people chosen from each strata (subgroup) is the same in both the sample and the population. For example, if 10% of an organization's members are classified managers, then 10% of the managers are randomly selected to be part of the overall sample. This approach yields a sample which is more representative of the population and will result in more precise needs analysis estimates.

If analysis of subgroups is a primary aim, it is important to make sure that subgroups are represented in reasonable numbers (i.e., 10 or more people in a stratum). In such cases, use ***disproportionate stratified random sampling***. This involves increasing the representation of smaller population subgroups while keeping the overall sample size the same. Because the proportion of sample cases in each stratum does not reflect its proportion in the total

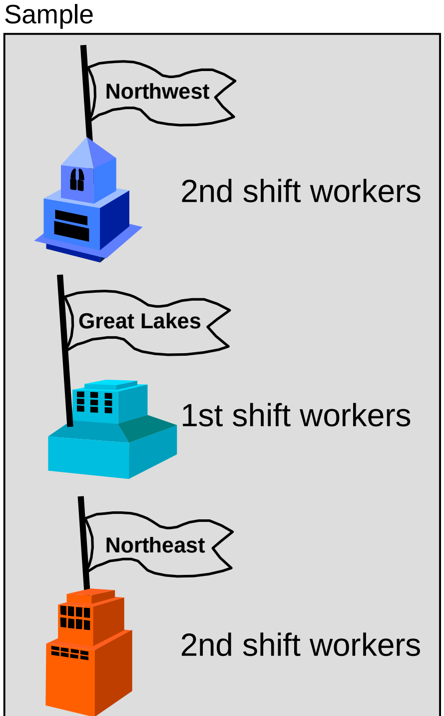
NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

Figure 2.7: Cluster Sampling (Natural Groupings)



Sample



1 Break the population into clusters of convenience:

For instance, a company's six regional operations located in various geographical areas of the country:

- Northwest Operations
- Southwest Operations
- Mountain Operations
- Great Lakes Operations
- Southeast Operations
- Northeast Operations

2 Place the names of the six regional operations in a hat and then select three of the six randomly.

3 Select employees at each of the randomly chosen sites.

a. Single Stage Example:
Use all of the workers from each of the chosen sites.

b. Multistage Example:
Randomly select from the cluster 1st, 2nd, or 3rd shift workers.

-- for each of the three randomly chosen sites, put the numbers 1, 2, and 3 in a hat, then select one number at random.

-- use all of the workers from the randomly selected shift for each site.

GUIDELINES

population, A WORD OF CAUTION! When cases are selected in this manner, a statistical adjustment must be made before generalizations from the performance analysis data are possible. This is accomplished by a weighting procedure that compensates for over sampling in some strata. These types of mathematical formulas are generally part of statistical analysis packages.

4. CLUSTER SAMPLING

In cluster sampling the population is broken down into groups of cases, called 'clusters', and a sample of clusters is selected at random. The clusters generally consist of natural groupings such as geographic units (e.g., districts, states, and regions) or work groups within or across organizational settings. Unlike stratified random sampling, which involves selecting cases from each stratum, cluster sampling selects cases only from those clusters selected for the sample.

Figure 2.7 illustrates cluster sampling.

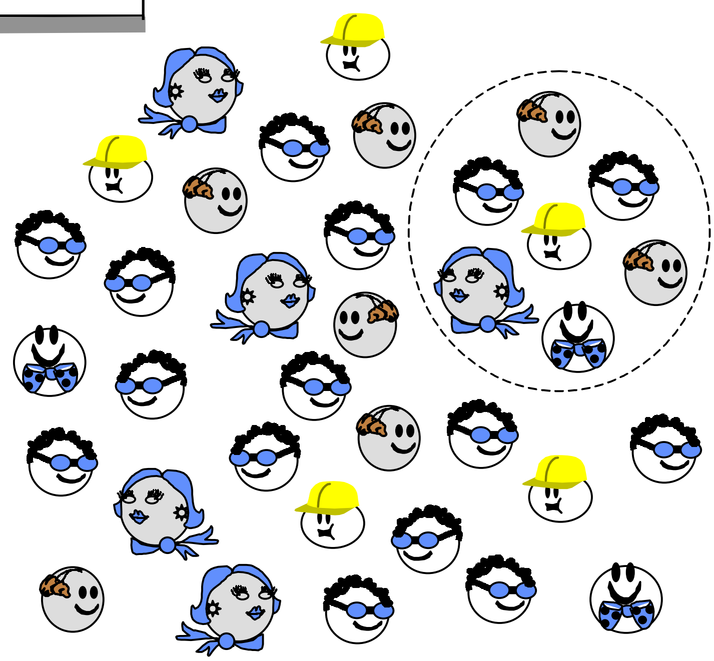
There are two types of cluster sampling designs: single-stage cluster sampling and multistage cluster sampling. If all cases in each sampled cluster are included in the final sample, the design is called a single-stage cluster sample in that sampling occurs only once -- at the cluster level. An example of a single-stage cluster sample would be to randomly select five of twenty work groups within an organizational work site and include all of the workers in each selected work group in the final sample.

However, cluster sampling more commonly involves sampling at two or more steps or stages; this is called multistage cluster sampling. An example of a two-stage cluster sample might involve comparing work groups across several organizational work sites. The first stage of cluster sampling would be to randomly select two of four organizational work sites; these would be termed 'primary sampling units'. The second stage of cluster sampling would be to randomly select five of twenty work groups within each of the selected organizational work sites; these would be termed 'secondary sampling units'. The final sample would include all of the workers within each selected work group within each selected organizational work site. You can also employ simple or stratified random sampling at one or more stages of the cluster sampling design, therefore, reducing the overall size of the final sample.

While stratified random sampling is used to increase sample precision (i.e., to adequately represent disproportionately small groups), the primary reason for cluster sampling is to reduce the time and costs of data collection. In other words, clustering concentrates data collection within fewer and smaller settings. In the example discussed in the previous paragraph, simple and stratified random sampling would involve randomly selecting and making contacts with individuals across all organizational work sites. Cluster sampling would lessen the number of sites and work groups ‘clustering’ the individuals to be contacted. The primary problem with cluster sampling, compared to simple and stratified random sampling, is that it yields a less representative sample, therefore limiting the generalizability of the research findings.

NOTES AND SUPPLEMENTAL INFORMATION

Figure 2.8: Convenience Sample: easy, but not always representative



The Sample

Seven individuals who are willing to take part in a focus group during their lunch hour.

- 1 of 3 Managers
- 1 of 5 Sales People
- 2 of 7 Office Support Workers
- 2 of 15 Production Workers
- 1 of 5 Delivery Drivers

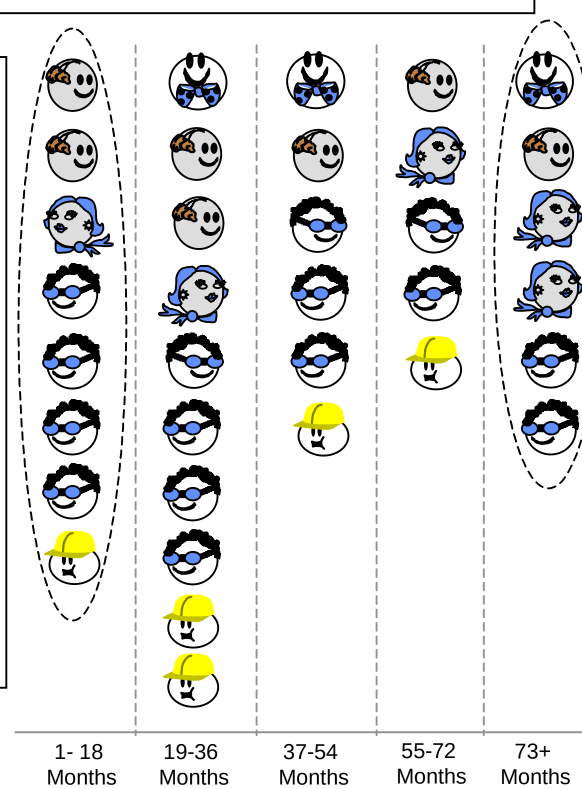
Figure 2.9: Purposive Sample

The Sample

1 It is deemed important to interview the most and the least experienced employees.

2 Group the employees by months on the job.

3 Select and interview all employees with 1-18 months and 73+ months of experience.



GUIDELINES

Non-Probability Sampling Designs

Non-probability sampling is oftentimes either more practical or the only viable means of case selection. The following are situations in which non-probability sampling would be the preferred method.

1. Under some circumstances, samples must include very few cases. If the population is quite large, then the selection of cases should probably be left to expert judgment rather than random chance. The rationale is that random selection of a few cases would not yield a representative sample.

2. When studying past events, analysts tend only to have access to a fraction of cases. For example, access to work samples or artifacts from a previous year might be quite limited, not allowing one to select cases randomly.

3. In the early stages of investigating a problem or a need, the objective is oftentimes to simply become more informed about the problem itself. Probability sampling is unnecessary in many such situations.

4. If the population itself contains few cases, then there is no point in considering probability or non-probability sampling. All cases (i.e., the population) should be studied.

1. CONVENIENCE SAMPLING

Figure 2.8 illustrates convenience sampling.

In this form of sampling, the analyst simply selects a requisite number from cases that are conveniently available. This type of case selection is easy, quick, and inexpensive. If the research is at an early stage and generalizability is not an issue, then this sampling method is appropriate. The weakness of this method, however, is that there is no way of determining to whom, other than the sample itself, the results apply.

2. PURPOSIVE SAMPLING

Figure 2.9 illustrates purposive sampling.

In this form of sampling, the performance analyst relies on his or her expert judgement to select units that are "representative" or "typical" of the population. The general strategy is to identify important sources of variation in the population and then to select a sample that reflects this variation.

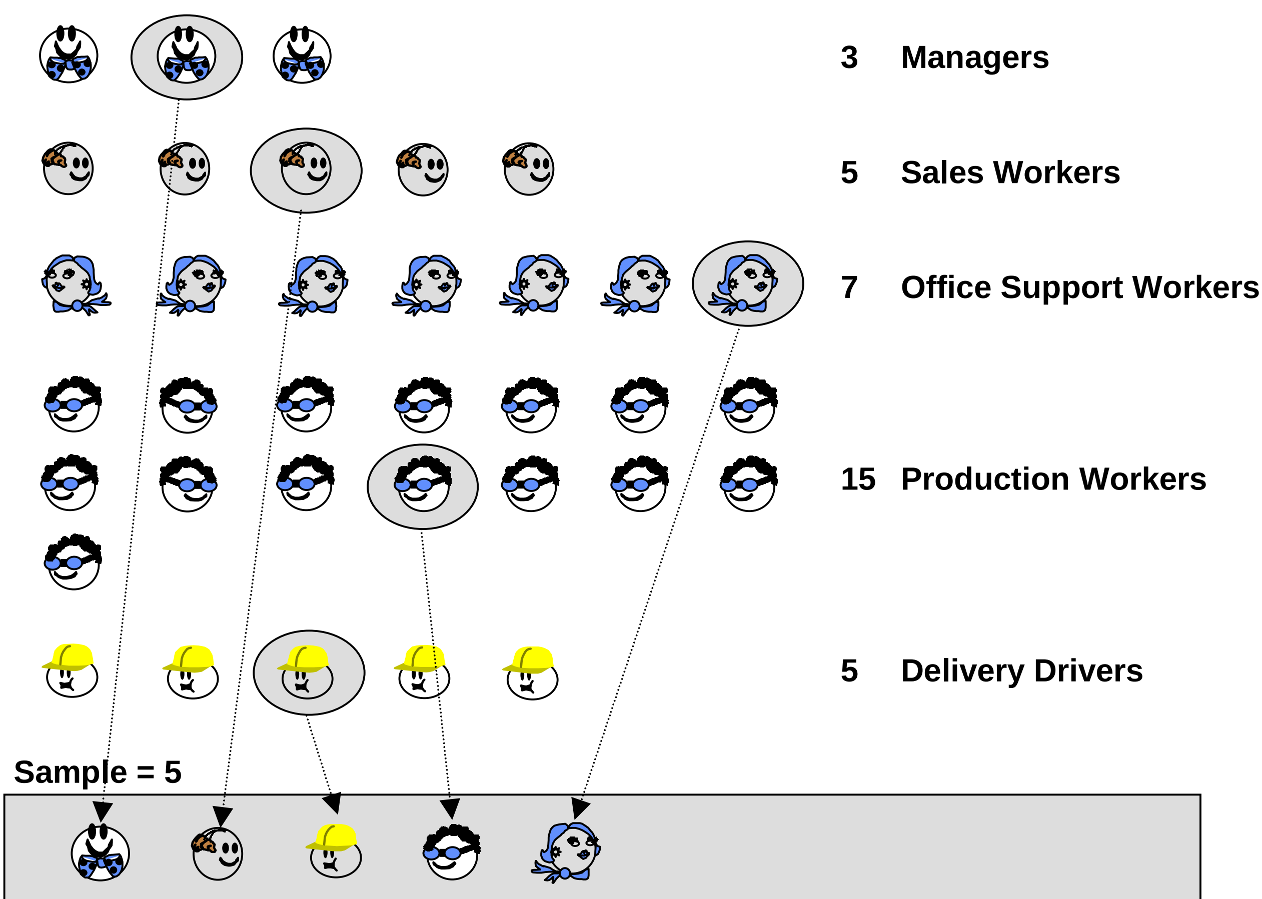
As with any non-probability case selection method, purposive sampling is an unacceptable substitute for probability sampling when precise and accurate generalizations are required. However, with studies of a more limited scope or in situations where random sampling is not possible, purposive sampling is an acceptable alternative. It generally offers much stronger, less tenuous

Figure 2.10: Quota Sample

The Sample

1 Divide the population into relevant strata.
Example: 5 workgroup types.

2 Use a non-probability sampling method to select cases from each strata.
Example: a convenience sample consisting of one volunteer from each of the five strata.



GUIDELINES

inferences than convenience sampling. The major weakness of purposive sampling is that it demands considerable knowledge of the population on the part of the performance analyst before the sample is drawn.

3. QUOTA SAMPLING

Figure 2.10 illustrates quota sampling.

Quota sampling is a form of purposive sampling that reflects the basic tenets of proportionate stratified random sampling. The first stage of case selection for both sampling methods entails dividing the population into relevant strata such as age, gender, race, or work group types. The difference between the two methods lies in how cases are selected once quotas have been set. Instead of randomly selecting cases within each stratum, quota sampling involves the selection of cases based on whatever non-probability method the performance analyst chooses (e.g., convenience or purposive sampling). Sampling biases are not completely eliminated using this method. However, it is considered to yield more representative samples than either convenience or purposive sampling.

Factors to Consider in Choosing a Sample Design

The following four questions must be considered before selecting a sampling design:

1. What is the stage of the performance analysis study?
2. How will the performance analysis data be used?
3. What are the available resources for drawing the sample?
4. How will the data be collected?

STAGE OF RESEARCH AND DATA USE:

The first two questions ask how accurate the sample must be as a description of the population. Accuracy is least important in the exploratory phases of performance analysis, when the goal is to simply discover patterns or to make initial observations. Under these circumstances, generalizing to a specified population is usually irrelevant. Therefore, non-probability sampling methods (i.e., convenience or purposive sampling) may be used to select cases.

In contrast, accuracy is most important in large-scale fact-finding performance analysis studies that provide input for major policy decisions. Probability sampling would be used in this type of situation (i.e., simple random, systematic, stratified, or cluster sampling).

In sum, the most important issue is that the performance analyst must have some sense of how representative the sample must be to meet the objectives of the performance analysis study.

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

Figure 2.11: Sampling Design Choice Factors

Factors Affecting Sampling Design

- **Stage of the study**
  - Early and exploratory → Non-probability
  - Fact-finding → Probability
  - Impacts policy → Probability
- **How are data to be used**
  - Developmental work → Non-probability
  - Finalized work → Probability
- **Available time**
  - Limited → Non-probability
  - Adequate → Probability
  - Plentiful → Probability
- **Available resources**
  - Limited → Non-probability
  - Adequate → Probability
  - Plentiful → Probability

How are Data Collected

- **Questionnaires** → **Probability**
- Structured Interviews → Probability
- **Unstructured Interviews** → Quota/Convenience
- Observations → **Purposive**
- **Artifacts - Population Accessible** → Probability
- Artifacts - Population Not Accessible → **Convenience**
- **Focus Groups** → Probability

GUIDELINES

AVAILABLE RESOURCES:

Sample accuracy must oftentimes be balanced against cost. Probability sampling generally involves more time, money, materials, and personnel in order for cases to be appropriately selected than does non-probability sampling. Similarly, comparing probability sampling techniques, simple random sampling is more costly than cluster random sampling. A general rule to follow is that the performance analyst should define the target population to fit the scope of the study. If the study is small enough, a well-designed sampling method is possible and will use limited resources.

METHOD OF DATA COLLECTION:

Figure 2.11 summarizes the factors that affect the choice of a sampling design.

The last factor to consider in selecting a sampling design is the method of data collection: surveys/questionnaires, interviews, observations, artifacts, or focus groups. The following are generally accepted sampling designs for the various methods of data collection.

SURVEYS AND INTERVIEWS: some form of probability sampling is most generally used with surveys and questionnaires. This also holds true for structured interviews (i.e., those conducted using structured instruments). For unstructured interviews, quota or convenience samples are often chosen.

OBSERVATIONS: Purposive sampling is typically used in performance analysis studies that use observations.

ARTIFACTS (e.g., DOCUMENTS): when the population is easily defined, probability sampling is used in the analysis of available objects (e.g., work samples or documents). However, if the population is not easily defined, convenience sampling is typically used.

FOCUS GROUPS: the accepted sampling design for focus groups is cluster sampling. Once clusters are chosen, it is best to use a form of probability sampling. If this is not feasible, purposive or quota sampling is advisable.

Determining Sample Size

After having identified the target population and decided upon an appropriate sampling design, the next step is to ascertain an adequate sample size. Five interrelated factors will affect the decision about sample size:

1. Heterogeneity of the population.
2. Desired precision of the sample.
3. Type of sampling design.
4. Available resources.
5. Number of breakdowns planned in the data analysis.

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

Table 2.3: Rules of Thumb* for Choosing Sampling Techniques and Sample Sizes for Different Types of Data Gathering Techniques Used in Needs Analysis Studies

Data Gathering Technique	Preferred Sampling Techniques	Suggested Sample Size
Document (Artifact) Analysis	1. Simple random 2. Systematic if artifacts are listed or organized in some fashioned. 3. Stratified random if subgroups of artifacts are important to the analysis	100 for each major subgroup, and 20 to 50 for minor sub-groups.
Observation	1. Purposive 2. Cluster if groups of people or objects are observed together. 3. Convenience if other options are not feasible.	Gather information until you reach a saturation point when newly collected data are redundant with data already collected.
Focus Group Interviews	1. Cluster with secondary units chosen randomly. 2. Cluster with secondary units chosen purposefully. 3. Quota if subgroups are well defined and understood.	7 to 10 people per group and 4 groups for each major audience type.
Unstructured Interviews	1. Quota 2. Purposive if subgroups are well defined and individuals identifiable. 3. Convenience if other options are not feasible.	Gather information until you reach a saturation point when newly collected data are redundant with data already collected.
Structures Interviews	1. Simple random 2. Systematic if it is easier to work with an established list of people. 3. Stratified random if subgroups of people are important to the analysis.	30 to 50
Questionnaires	1. Simple random 2. Systematic if it is easier to work with an established list of people. 3. Stratified random if subgroups of people are important to the analysis.	100 for each major subgroup, and 20 to 50 for minor subgroups.

* For more details and precise guidelines, see Chapter 10 in Mertens, D. M. (1998). Research methods in education and psychology: Integrating diversity with quantitative & qualitative approaches. Thousand Oaks, CA: Sage.

GUIDELINES

Table 2.3 shows some general rules for choosing sampling techniques and sample sizes.

POPULATION HETEROGENEITY: Heterogeneity refers to the degree of dissimilarity among cases with respect to a particular characteristic while homogeneity refers to the degree of similarity. The more heterogeneous the population with respect to the characteristic being studied, the more cases required to yield a representative sample.

DESIRED PRECISION: Desired precision refers to the extent to which the findings collected from the sample would reflect the findings if the entire population were to be studied. In general, the larger the sample, the more precise the findings. In fact, the absolute sample size is a better determinant of precision than the proportion of the population sampled.

SAMPLING DESIGN: The type of sampling design also affects decisions about sample size. For example, one way of increasing precision other than selecting a larger sample is to use stratified rather than simple random sampling. In other words, stratified random samples require fewer cases than simple random samples. In contrast, cluster sample sizes should be larger than simple random samples in order to achieve the same degree of precision.

AVAILABLE RESOURCES: Sample size, sampling method, and method of data collection all impact the cost of implementing a performance analysis study. The following rule of thumb is typically used in survey research.

Once the data collection method is specified, the sample size can be determined on a time and cost per case basis. In other words, the number of cases will be equal to the total time for data collection divided by the time per case, OR the total funds for data collection divided by the cost per case.

NUMBER OF BREAKDOWNS PLANNED: The number of variables and categories into which the data are to be grouped and analyzed must also be taken into account in determining sample size. The more breakdowns planned in the analysis, the larger the sample must be.

For example, let's assume that we have a sample of 1,000. We divide the sample into males/females, then into African-Americans/whites, then into people working the first or second shift, and then to those working in eastern and mid-western plants. When we look at the number of mid-western white females working the second shift, we might find that the sub-sample of this group is too small for a reliable analysis of this particular group.

Example Sample	1000
Females	289
White	60
Working second shift	10
Mid-western plants	4

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

Table 2.4: Minimum Sample Size for a ±5% Sampling Error with a 95% Confidence Level

Population Size	Sample Size	Percent Required
10	10	100
20	19	95
50	44	88
100	80	80
250	152	61
500	217	43
1,000	278	28
2,500	333	13
5,000	350	7
10,000	370	4

From Krejcie, R.V., and Morgan, D. E. (1970). Determining sample size for research activities. Educational and Psychological Measurement, 30(3), 607-610.

Table 2.5: Margin of Error for Sample Sizes With a Confidence Level of 95%

Sample Size	Number of percentage point to allow
100	7 to 11
200	5 to 8
400	4 to 6
600	3 to 5
750	3 to 4
1,000	2 to 4

The Gallup Organization, publisher of many polls, and *The Gallup Poll Monthly*, uses the guidelines in Table 9.5 to estimate the 95% margin or error. From Patten, M. L. (1998) Questionnaire research: A practical guide. Los Angles, CA, Pyrczak.

GUIDELINES

To avoid such problems, it is important to estimate the number of breakdowns prior to data analysis and to make sure that the total sample size will provide enough cases in each subcategory. One way to achieve this and also keep the sample size relatively small would be to employ stratified or quota sampling.

Sampling Error and Sample Size

When a probability sample is used, the results may be close to the actual population values, but they will NOT be exactly the same. This error is called ***sampling error***. It occurs because each sample drawn from a population will be different from other samples that can be drawn and all samples will be different from the population. Sampling error is influenced by two factors: 1) the amount of variability in whatever is being measured -- the more variability, the higher the error, and 2) the size of the sample -- the larger the sample, the lower the error.

The ***sampling error*** is sometimes called the ***margin of error***. The two terms mean the same thing.

A specified ***degree of confidence*** is also called a ***confidence interval***.

When a sample statistic such as the percentage of persons responding in a particular fashion is calculated, it is only an estimate of the percentage of respondents in the population. As such, sample statistics, such as percentages, are often reported in conjunction with a sampling error. For many polling data reported in the popular press, the sampling error is reported as ∀ 3% or ∀ 5%. For instance, if 75% of workers desire more flextime, and the sampling error is ∀ 5%, this means that it is very likely that the population percentage of workers desiring flextime falls between 70% and 80% (that is, 75% ∀ 5%).

Note in the paragraph above, that the words “very likely” were used in explaining a specified interval. It was noted that a true value can only be known if the entire population is measured. Statistically, we can state a ***degree of confidence*** for the results. Commonly used degrees of confidence are 90%, 95%, and 99%. For the example above, we could stipulate that the true population value would be between 70% and 80% in 95 out of 100 randomly drawn samples.

Tables 2.4 and 2.5 present guidance for sample sizes for various sampling errors and degrees of confidence.

The desired sampling error and degree of confidence need to be considered when deciding on an appropriate sample size for a probability sample. Various formulas exist to calculate the required sample size. They are present in most basic statistics books and built into many statistical software packages. The tables to the left provide some guidance for minimal sample sizes given specific degrees of confidence and sampling errors.

Minimum and Actual Sample Size

As stated above, research methods and statistics books abound with formulas and tables which provide guidance on sample sizes given specific confidence levels and sampling errors. Note however, that most will be giving ***minimum*** sample sizes. The assumption is that every person, object, or event sampled will be available for analysis. We know that in most

GUIDELINES

NOTES AND SUPPLEMENTAL INFORMATION

Figure 2.12: What Sample Size to Use?

A **conservative algebraic formula** for calculating minimal sample size for a given margin of error at the 95 percent confidence interval.

$$\eta = \frac{1}{e^2}$$

where η = sample size
e = margin of error

EXAMPLE:

For a 3% margin of error: $\eta = \frac{1}{e^2} = \frac{1}{(.03)^2} = \frac{1}{.0009} = 1{,}111$

Example:

Margin of Error (e)	Sample Size
1 %	10,000
2 %	2,500
3%	1,111
4%	625
5%	400

Examples of values that have been calculated using the conservative algebraic formula presented above.

A **formula that simultaneously corrects** for ineligibles and non-responses.

$$\eta' = \frac{\eta}{(e)(r)}$$

where η = the minimum sample size
e = proportion of eligible respondents
r = expected response rate

EXAMPLE:

XYZ corporation has 2,550 production workers. The analysis team wants to survey the workers using a simple random sample. For a 5% margin of error and 95% level of confidence, a minimum of :

$$\eta = \frac{1}{(e)^2} = \frac{1}{(.05)(.05)} = \frac{1}{.0025} = \mathbf{400}$$

The analysis team estimates that 90% of the target population will be available to complete the survey and the team expects a response rate of 80%.

$$\eta' \textbf{ (adjusted sample size)} = \frac{\eta}{(e)(r)} = \frac{400}{(.90)(.80)} = \frac{\mathbf{400}}{\mathbf{.72}} \textbf{ or } \mathbf{555.5} \textbf{ or } \mathbf{556}$$

Therefore, 556 surveys should be administered to obtain the necessary minimum sample size.

GUIDELINES

Figure 2.12 presents formulas for choosing a sample size and for correcting for ineligibles and non-respondents.

It is important to note that after a certain point, sample size does not increase in proportion to population size. In table 2.4 the sample size for a population of 5,000 is 350 and for a population of 10,000 only 370.

Performance analysis studies, this will not be the case. Some people, objects, or events will not be available and some people may decline to participate. These ineligibles and non-respondents require that the sample size be larger than the minimum. Henry in his 1990 book titled *Practical Sampling* provides a simple formula that simultaneously corrects for ineligibles and non-response. The formula and an example are given on the page to the left.

Final Thoughts On Populations and Samples

For most performance analysis projects, studying a population is preferable to a sample. When using questionnaires or analyzing artifacts and the population size is 250 or below, use the population. For larger populations, use a sample. Mostly use simple random sampling. If particular dimensions of the population are deemed important to the study, use a stratified random sample. A 95% confidence level and a margin of error of ∀ 5 are conventional for performance analysis type studies. When people are the unit of analysis, always sample non-respondents to see if there is any bias between those who choose to participate and those who do not.

For exploratory and developmental performance analysis work, consider non-probability sampling. Be precise in choosing the unit of analysis for a study and be careful in generalizing beyond both the unit and from a non-probability sample to a population. Time, resource expenditures, and financial costs are all considerations in conducting a performance analysis study. However, it is better to error on the side of a population study or a larger sample if the findings need to be credible and defended against “nay sayers.” In short, take the time and resources to increase a sample. Even consider seeking help from a technical consultant in setting up a study, choosing a sample, and analyzing the data.

USCG

WORKSHOP

S
U
R
V
E
Y

Section 3: Graphing Data

CONTENTS

Overview

PAGE

Guidelines

PAGE

Legend:

Reference notes and supplemental information on opposite page.

References to parts of the handbook and other ideas.

1 Numbered steps and Procedures

Important points, hints and ideas.

Advantages

Limitations

Key Points and Ideas

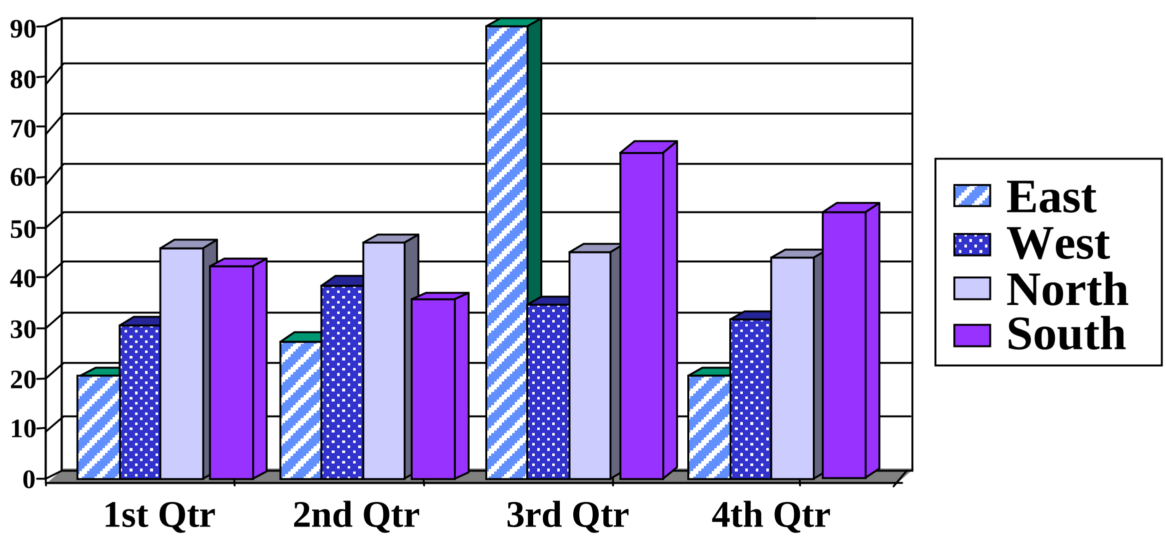
NOTES AND SUPPLEMENTAL INFORMATION

OVERVIEW

Figure 3.1: Visual Illustration of the Advantages of Graphically Displaying a Set of Data

Quarterly Sales in Thousands of Dollars

	1st Qtr	2nd Qtr	3rd Qtr	4th Qtr
East	20.4	27.4	90.0	20.4
West	30.6	38.6	34.6	31.6
North	45.9	46.9	45.0	43.9
South	42.3	38.9	66.0	53.2



Key Reasons for Drawing a Chart or Graph

- Charts and graphs give immediate depictions of the differences and patterns in a set of data
- A lot of data can be seen quickly
- Charts and graphs create pictures in the readers' mind

OVERVIEW

Introduction

After data have been gathered for a performance analysis study, it is often helpful to organize and display the data in charts, graphs, and tables. This will enable the analyst and others interested in the results of the study to better grasp the meanings of the data, to draw conclusions, and to make recommendations.

When graphing a specific set of data, it is important to pay attention to three major issues:

- selecting the chart, graph, or table that is most appropriate to best showcase the data being illustrated
- understanding the levels of interests and abilities of the readers to decode the graphic
- following conventional graphics production rules in order to help the reader interpret the graphic easily and accurately

Figure 3.1 illustrates the advantages of graphically displaying a set of data.

Graphically displaying data can be very effective in quickly presenting a picture in the reader's mind. However, as Edward Tafte (1983) points out, it is important to remember that a statistical graphic should let the reader see what is happening over and above what has already been described in the text of a report. If you think a report just needs some color or pictures to break up the text, then choose something other than statistical graphics.

Purposes of Graphing Data

Graphs and charts show the differences and similarities in the values of numerical data without having to compare the entire set of numbers.

There are a number of good reasons for choosing to use a visual representation of needs analysis data:

- to orient readers to the subject or topic
- to convey a message or reinforce a central point
- to point out a problem or issue central to the analysis
- to provide an overview of the issue or opportunity being studied
- to allow readers to generate questions and issues for themselves
- to expose a commonly held myth or misconception about the problem or opportunity being studied

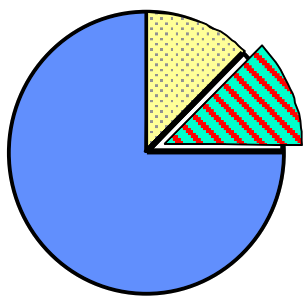
Usually, the way an analyst chooses to display the performance analysis data will be based on three factors:

❶ What is it that the analyst is trying to explain with the data
❷ Who will be looking at the chart, graph, or table
❸ The analyst's understanding and skill in producing different types of graphical displays of data

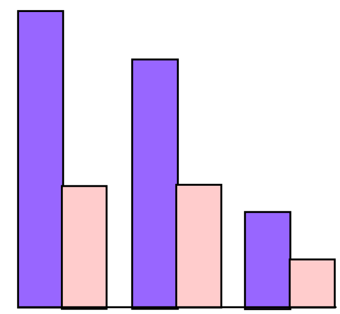
NOTES AND SUPPLEMENTAL INFORMATION

OVERVIEW

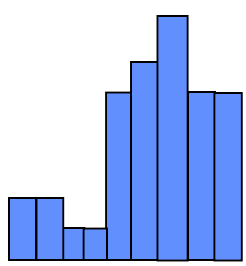
Figure 3.2: Types of Charts and Graphs



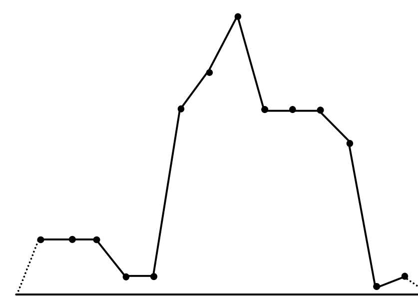
Pie Chart



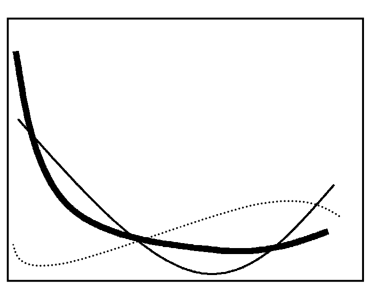
Bar Chart



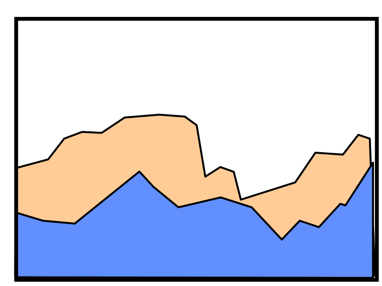
Histogram



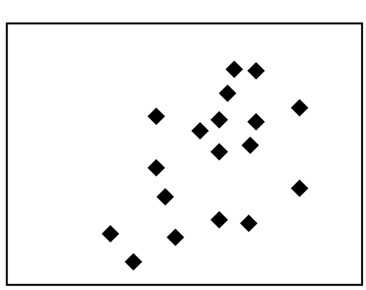
Frequency Ploygon



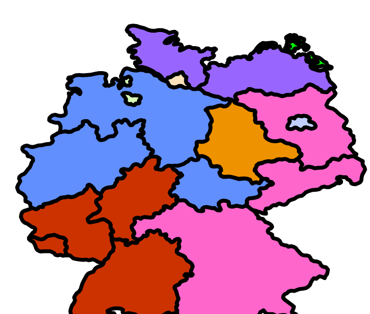
Line Graph



Area Chart



Scattergram



Choropleth Map

OVERVIEW

Considerations in Creating Charts and Graphs

Creating charts and graphs is part craft and part art. The craft involves developing a technical understanding of the basic elements of charts and graphs and knowing how to draw them. The art involves choosing the right graph or chart for the right situation.

There are two major determinations that must be made prior to choosing a particular type of chart or graph to use to visually display data.

❶ **TIME:** Were the measures taken at one or several points in time? If they were taken at one point in time, these are called **cross sectional data**. If they were taken at several points over a time period, they are called **time series data**.

❷ **TYPE OF DATA (VARIABLES):** Are the data which were collected qualitative, quantitative, or a combination of both?

Qualitative Data are categorical in nature. They are labels used to identify or name variables. Such data are called **nominal**.

Quantitative Data are numerical in nature. If they are discrete, they have only a few categories usually consisting of whole numbers. Such data are called **ordinal**. If the data are continuous, there are a large number of measures that are often infinite. Such data are called **interval** or **ratio** data.

The types of data we are dealing with influences the type of chart or graph we choose to display those data. The chart or graph must retain the characteristics of the data.

These distinctions are important because they influence the type of chart or graph that is appropriate to visually display a given set of data. For more detail about the distinctions between **nominal**, **ordinal**, **interval**, and **ratio data**, **see Section 4** titled "Descriptive Statistics."

Different Types of Charts and Graphs

Charts and graphs take many forms. Some are straightforward and simple while others are intricate and complex. Some are very specific and used for only one type of data and others can be used for more than one type of data. Anders Wallgren and his colleagues in their monograph titled **Graphing Statistics & Data** (1996) present over 25 different types of charts and graphs.

Figure 3.2 shows the basic structures of the eight types of charts or graphs dealt with in this section.

This section will look at eight basic types of charts or graphs which represent those used most often for visually displaying performance analysis data. The types are: **pie charts**, **bar charts**, **histograms**, **frequency polygons**, **line graphs**, **area charts**, **scattergrams**, and **choropleth maps**.

NOTES AND SUPPLEMENTAL INFORMATION

OVERVIEW

Table 3.1: Characteristics of Selected Charts and Graphs

Pie Chart	Used to show percentage or proportional distributions of categorical (nominal) data.	OK if the categories are few in number and it is not important for the reader to determine the precise size of the sectors with the eye.
Bar Chart	A basic chart used to illustrate distinct categorical (nominal) data. The bars can be vertical or horizontal.	Simple to draw and simple to read. Bar charts use spaces between the bars to emphasize that the data are discrete. They can be used to show both absolute and relative frequencies.
Histogram	Use to show frequencies for a continuous variable (interval or ratio data). The bars are vertical.	Histograms can show absolute or relative frequencies. There is no space between the bars to indicate that the data are continuous. In principle, the area of a bar is proportional to the frequency (= base times height).
Frequency Polygon	An alternative to a histogram when showing a continuous variable (interval or ratio data). Good for smoothing out the look of the data.	Constructed by connecting what would be the mid-points of the histogram columns with lines. Useful for comparing more than one set of continuous data on the same chart.
Line Graph	Good for showing data over time. The gradients of the curves can indicate periods of time when changes were large or important turning points. Used with continuous variables (interval or ratio data).	The line chart gives a clear picture of development. The eye can follow the curves over time and see important changes. A line graph can handle any number of points in time and is suitable when there are several series of data to compare.
Area Chart	Built upon a line chart, an area chart divides the data into two or more parts. The total is described by the top line and parts by lower lines. The different areas are colored or shaded.	Useful for comparing the whole data with distinctive parts of the data. It is best to keep the number of parts limited. Darker color or shading is used at the bottom.
Scattergram	Use to show relationships. The data consists of pairs of quantitative variables.	The pairs of data are plotted on a pair of coordinates (x & y). Each pair is represented by a dot or circle. The data can represent covariance or a causal relationship.
Choropleth Map	A shaded map which shows proportions, intensities, or averages for areas.	Classes or categories on the map's surface are distinguished using color, shading, or patterns. They are best used to illustrate ratios.

OVERVIEW

PREPARING TO DRAW A CHART OR GRAPH

Construction Rules And Guidelines

- Who will be looking at the chart?
- What does the data represent?
  - ✓ subject- is design information-rich graphics helpful?
  - ✓ target audience - is my graphic too difficult?
  - ✓ give primacy to the data
- Is a chart necessary, or is a table or descriptive text more suitable?
- What chart or graphic is appropriate given the variables?
- If the data are complex, split into several separate, smaller charts/graphs.
- Must the chart or graph have color?
- What is the final product? -- overhead, on computer, wall chart, handouts?
- Plan designs to reveal data clearly.
  - ✓ does it avoid distortion
  - ✓ does it enhance clarity (length, size, volume, labels, titles, legends)
  - ✓ does it encourage important comparisons
- Sizing the graph (general rules):
  - ✓ 1.2 - 2.2 wider than tall
  - ✓ 5” wide graph (on 81/2” x 11” paper) 2 1/4” to 2 1/2” tall
- Aesthetic appeal.
  - ✓ legibility
  - ✓ method of reproduction
  - ✓ placement on page

Table 3.1 describes the characteristics of selected charts and graphs.

Further Readings

Holcomb, Z.C. (1998). Fundamentals of descriptive statistics. Los Angeles, CA: Pyrczak Publishing.

Jones, J.E. and Bearley, W.L. (1995). Surveying employees. Amherst, MA: HRD Press.

Wainer, H. (1997). Visual revelations. New York, NY: Copernicus.

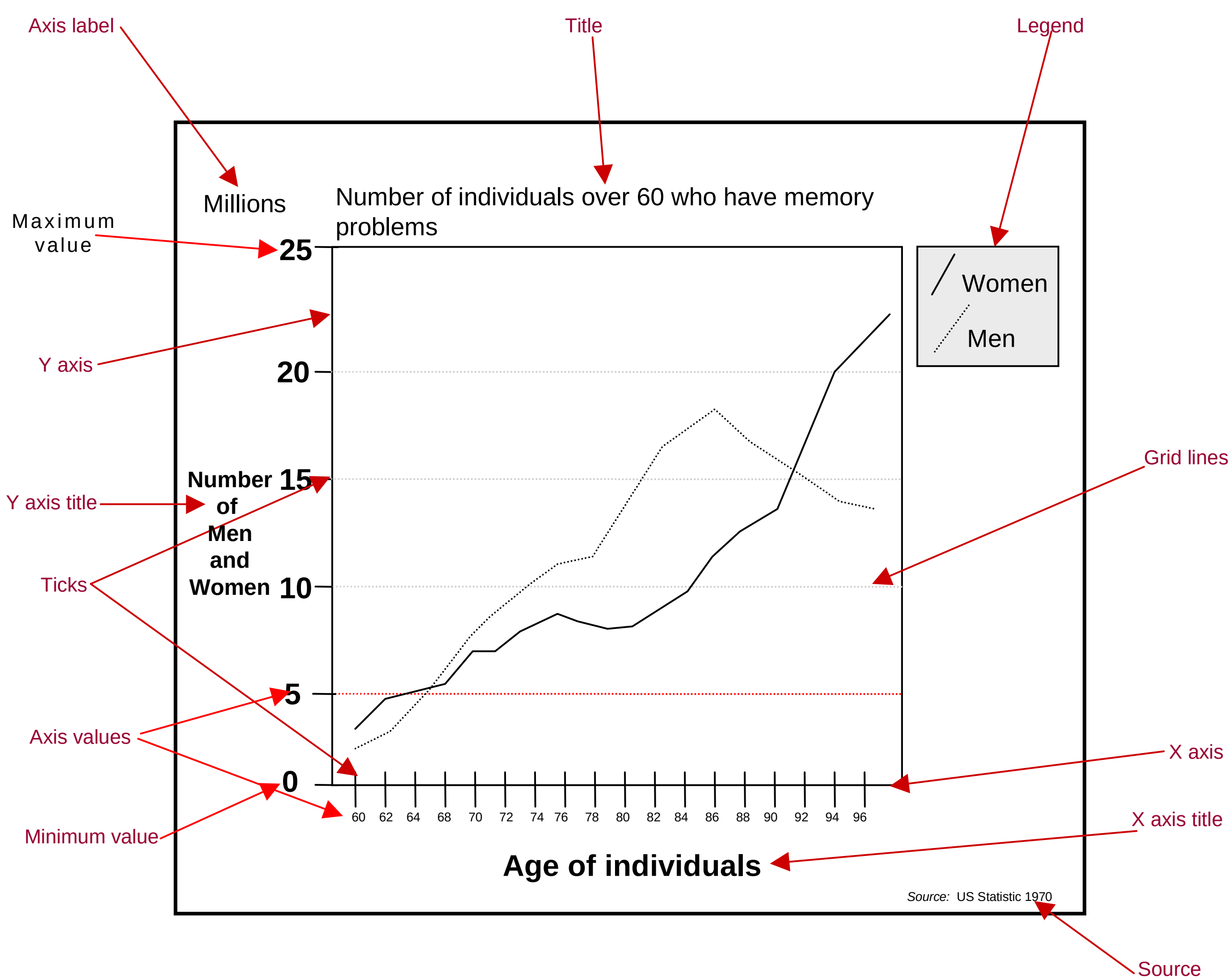
Wallgren, A., Wallgren, B., Persson, R., Jorner, U. and Haaland, J. (1996). Graphing statistics & data. Newbury Park, CA: Sage.

Figure 3.3: The Parts of Charts and Graphs

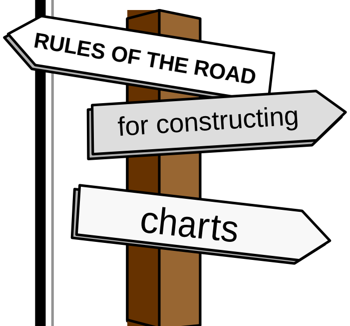
The parts that make up a chart and graph include:

Identifying Information
- Title
- X Axis Title/Label
- Y Axis Title/Label
- Source
- Legend

Parts
- X Axis
- Y Axis
- Grid Lines
- Ticks
- Axis Values
- Minimum Value



GUIDELINES



General Guidelines for Constructing Charts and Graphs

Figure 3.3 shows the key parts of charts and graphs.

CHART AREA Horizontal is best 1:1.6 height:width

BACKGROUND Light shading; if color use light background

FRAME/GRID Helpful; keep light and proportional

TEXT Keep to a minimum; words should not dominate
Title at top is best
- describe in concise terms what is shown
- left justified when part of text
- centered when self-standing

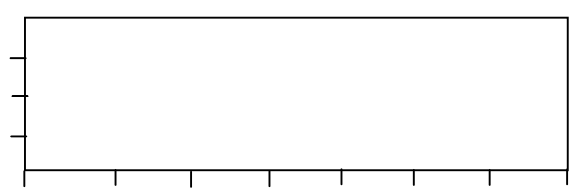
All text should be horizontal; left to right if designing for western culture

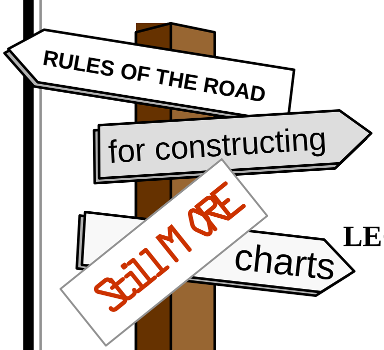
AXES Straight lines; do not need arrows
Y axes, vertical; text above or to left
X axes horizontal; text below or to right

SCALES Avoid big numbers; if thousands, millions, etc., note on a scale; use 1, 2, … 5 principle

0 1 2 3 4 5
0 2 4 6 8 10
0 5 10 20 30 40 50
0 25 50 75 100 for percentages
0 .1 .2 .3 .4 .5

TICKS Draw ticks on the outside of the axes





LEGENDS If space allows, write directly on the chart or graph
If space is small, write plot next to respective area
If there are several bars, stacked bars, or several lines, use a legend box

PATTERNS Use simple patterns, dots or lines 4 or 5 maximum
Darker patterns down or to the right

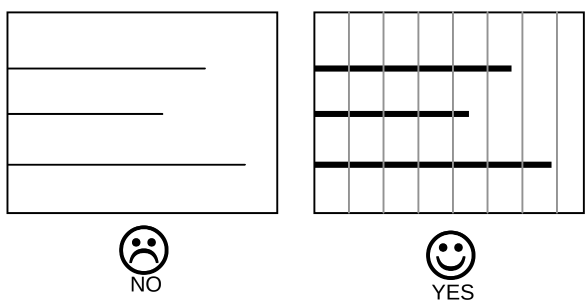
SHADING Must be distinguishable; darker shades down or to the right

COLOR Use sparingly; discreet colors against pale shading

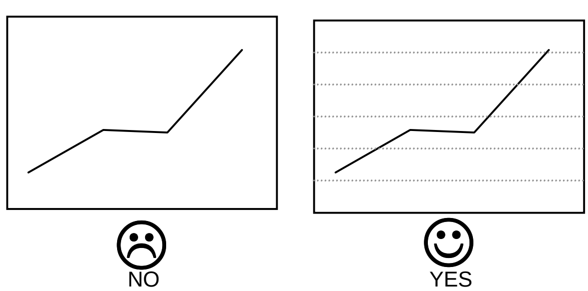
NOTES AND SUPPLEMENTAL INFORMATION

OVERVIEW

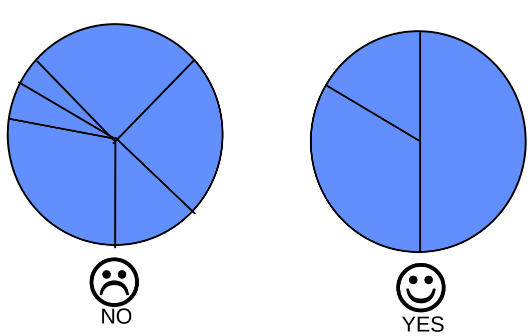
Figure 3.4: Helping the Reader Visually



Grid lines help the eye to judge where end points actually end.



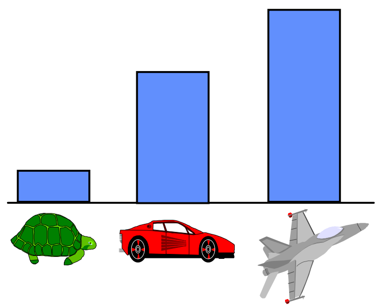
Grid lines are very subtle and still effective in showing the values at different points on a line graph.



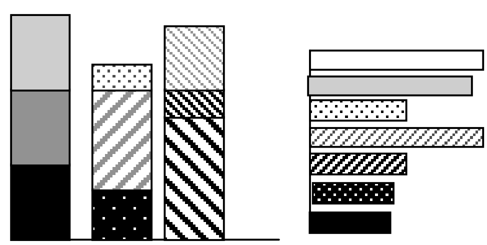
Estimating the size of the areas of a pie chart is difficult when the size of the areas are small or nearly equal.

GUIDELINES

Things to Do in Constructing Charts and Graphs

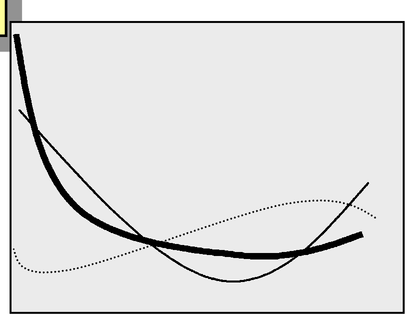


While pictographs are discouraged (i.e., using pictures in stacks to depict data), using easily understood symbols as categorical labels assists readers to grasp content quickly. Especially if the readers do not all speak the same language or categories are complex or confusing.

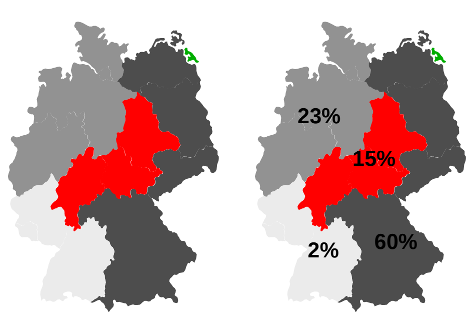


Choose patterns, colors, or shades which are different. Always put the darkest pattern or color at the bottom.

Figure 3.4 shows guidelines for visually helping readers.



Color can help draw the readers' eye to an important part of a graph. But be careful. If it is to be printed in black & white, use contrasting colors or patterns which will be distinguishable in black & white.

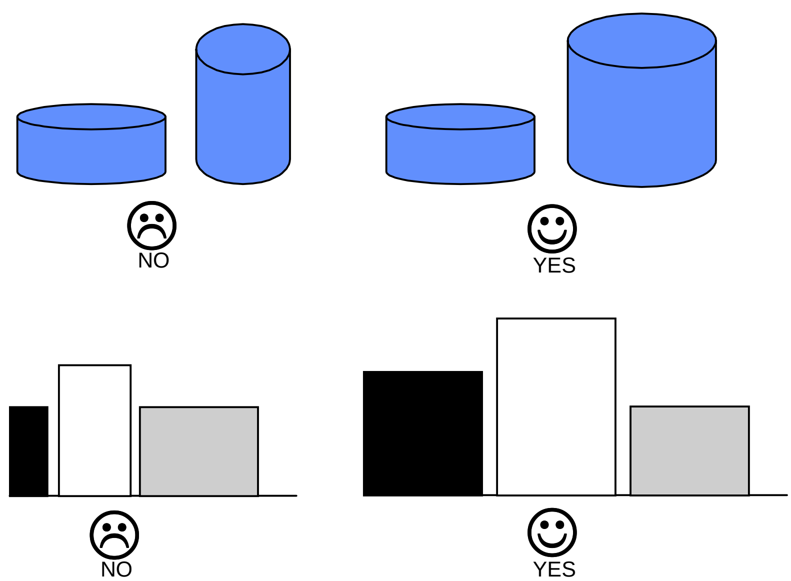


Distinguishing between shadings is very difficult. Use something in each area to distinguish between the areas.

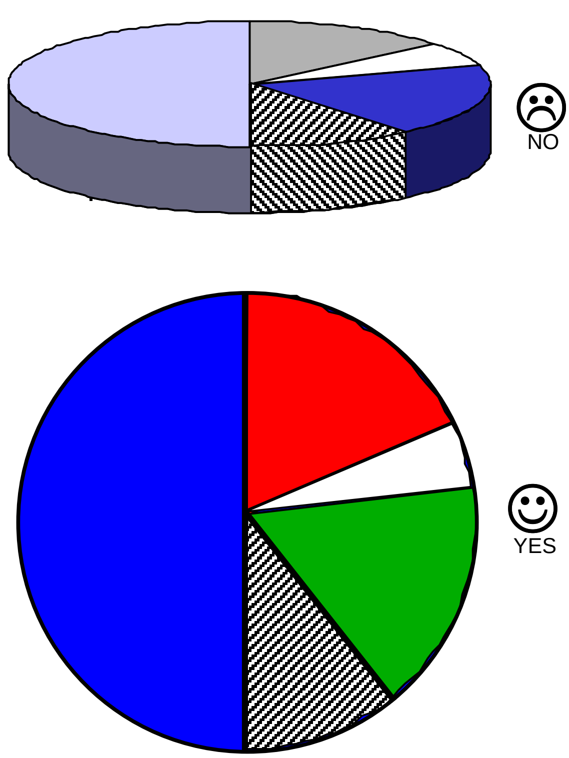
NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

Figure 3.4 (continued): Helping the Reader Visually



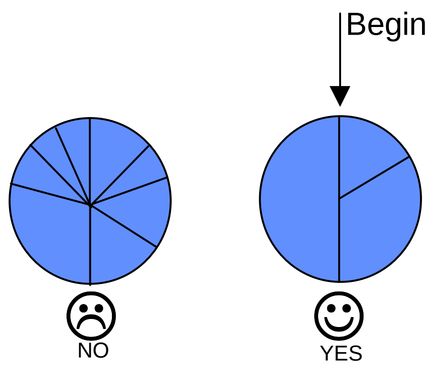
Illustrating quantitative estimates with volumes, even when using shading, is almost impossible. Use the shapes with the same widths to depict varying sizes.



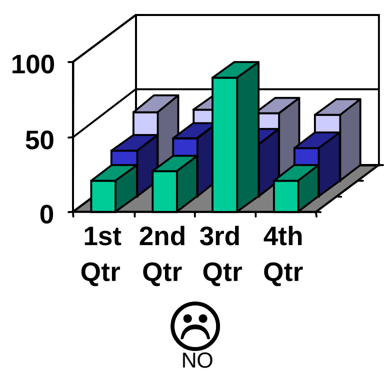
With computer technology it is easy to make three dimensional charts. However, when conveying two dimensional data, three dimensional charts distort the data.

GUIDELINES

Things to be Aware of in Constructing Charts and Graphs

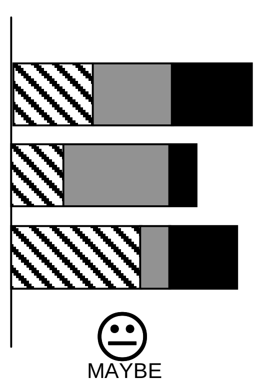


A reader must be able to judge areas and angles to use a pie chart. Don't use more than five or six sectors. If you need a category for 'other', make sure it is not dominant and place it at the end. Pie charts are read clockwise starting at 12:00.

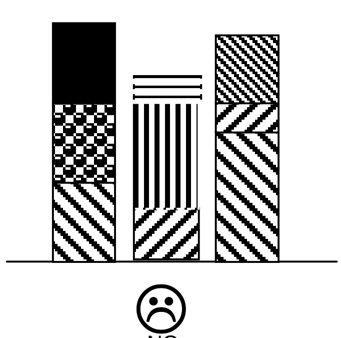


Beware of 3-D. Computers can make it easy to construct such graphics, however, they are difficult to interpret and it can skew the real comparisons being made.

Figure 3.4 (continued) shows additional guidelines for visually helping readers.



Stacked bars, while not the best, only show comparisons of lengths and positions on the scale. The actual value of each part is difficult to interpret.

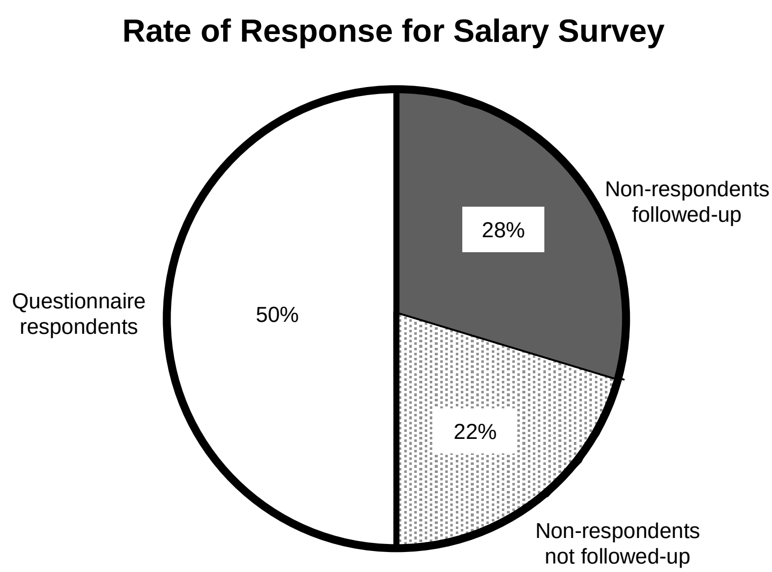


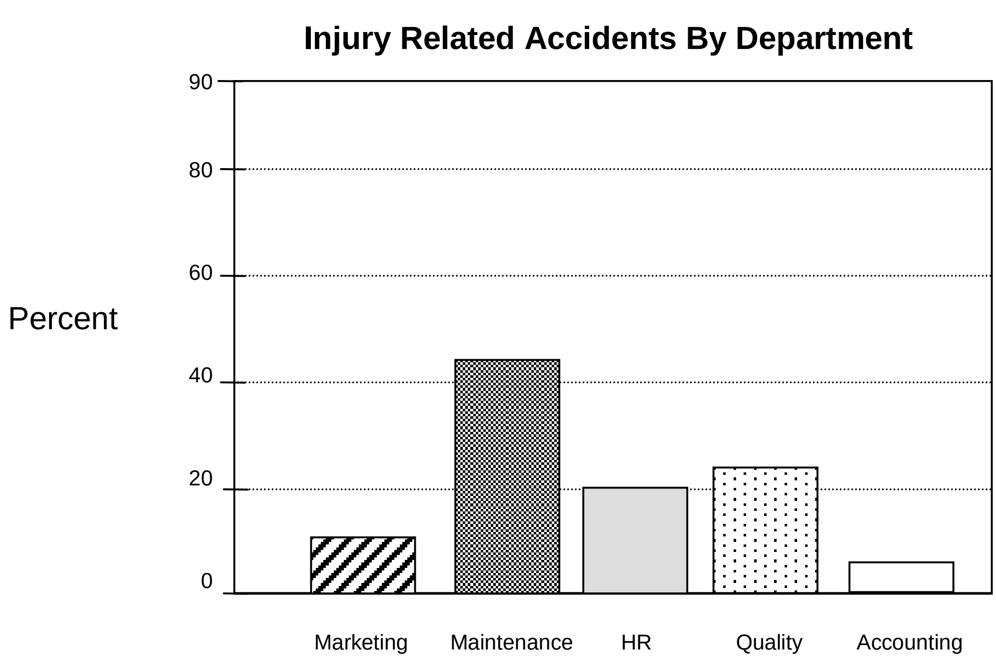
Avoid placing dark colors at the top of bars in a chart and mixing horizontal and vertical patterns. They are "busy" and confuse the reader.

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

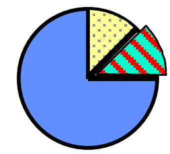
Figure 3.5: Example Pie Chart and Bar Chart





GUIDELINES

Constructing a Pie Chart



Type of Data:
- Nominal or categorical

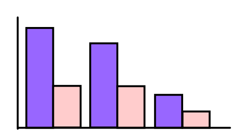
Uses:
- Comparing proportions
- To provide a general picture or overview of the situation

Guidelines:
- Begin at “12:00” and read clockwise
- Area = 100%
- 5 or 6 sectors are ideal
- Use different charts when comparing several groups

Cautions:
- Difficult to determine precise size of sections but relative size is easy to judge
- Keep flat or on one dimension. Three dimensional pie charts give misleading comparisons
- Make certain that the reader is able to judge the angles and areas
- Place numerical values in the diagram since there is no scale axis
- If using different colors and patterns, go from dark to light
- Seldom any reason to “explode” more than one sector

Figure 3.5 shows an example of a completed pie and bar chart.

Constructing a Bar Chart



Type of Data:
- Nominal or categorical

Uses:
- Simple to draw and read
- Use absolute or relative (%) frequencies

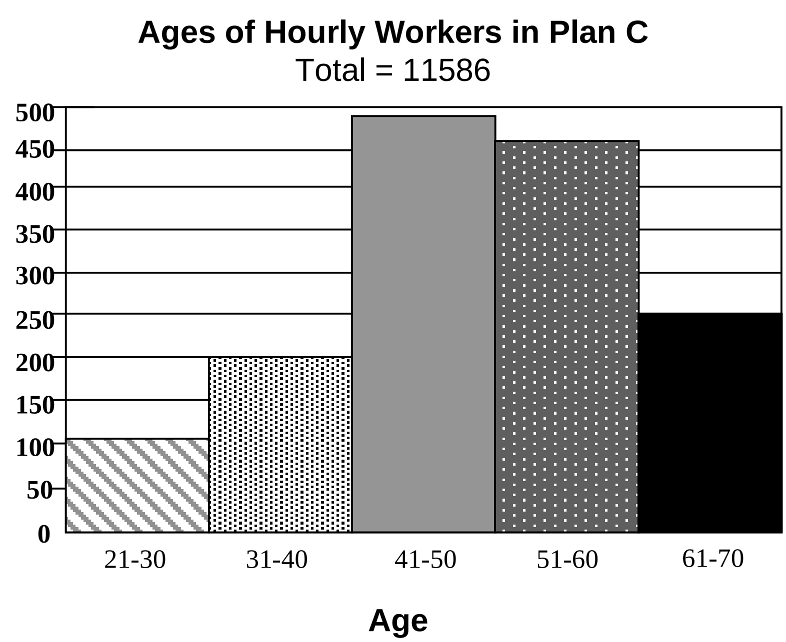
Guidelines:
- Space between bars indicate discreteness
- Gaps between bars; bars wider than gaps
- Grid lines help with comparisons
- Place horizontal when variable value labels are long or there are many variables
- Order of bars; go long to short (top down or left right)
- Grouped bar charts; only show two or three categories
- Overlapping bars; put categories with shortest bars up front in lighter shading

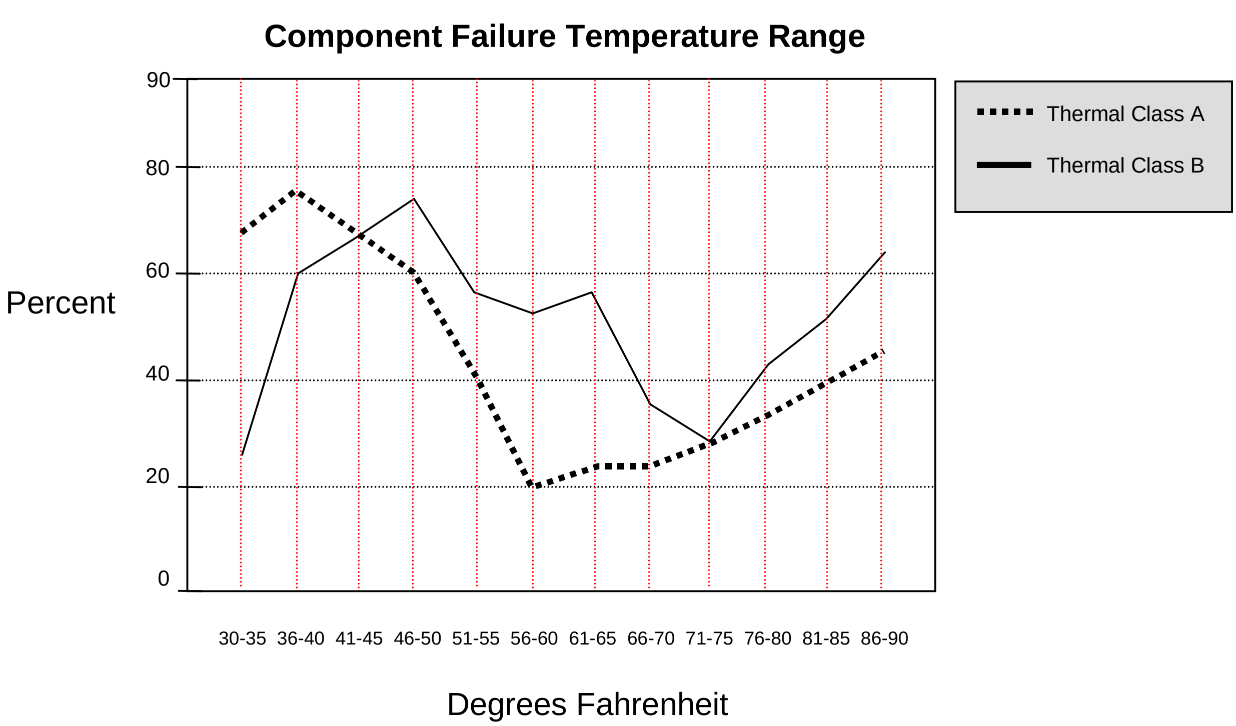
Cautions:
- Use stacked bar charts sparingly. The size of the categories are difficult to read
- Overlapping bar charts work only when a category has shorter bars throughout

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

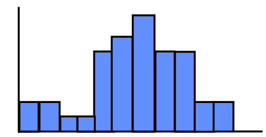
Figure 3.6: Example Histogram and Frequency Polygon





GUIDELINES

Constructing a Histogram



Type of Data: ✓Continuous (interval or ratio)

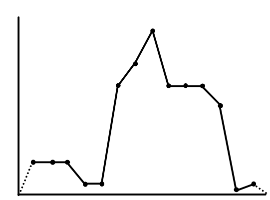
Uses: ✓Can show absolute values or relative frequencies (percentages)

Guidelines:
- ✓Place data in a table first for easy reference
- ✓If data points are numerous, compress into larger classes
- ✓X-axis represents class boundaries or name of class
- ✓With classes of equal size, the base of each bar is equal width
- ✓If classes are different in size, base of each bar should be proportional
- ✓In some situations it is OK to use open-ended classes -- when not sure of beginning or ending or there is none

Cautions:
- ✓If you are comparing different subgroups, use more than one histogram rather than a stacked histogram
- ✓If the variation is limited, it is difficult for the reader to see the size of the differences

Figure 3.6 shows an example of a completed histogram and frequency polygon.

Constructing a Frequency Polygon



Type of Data: ✓ Continuous (interval or ratio)

Uses:
- ✓ Alternative to a histogram
- ✓ Smoothes out the look of a histogram when the bars change dramatically
- ✓ Useful in comparing two or three variables; in place of grouped bar charts
- ✓ Can represent percentages

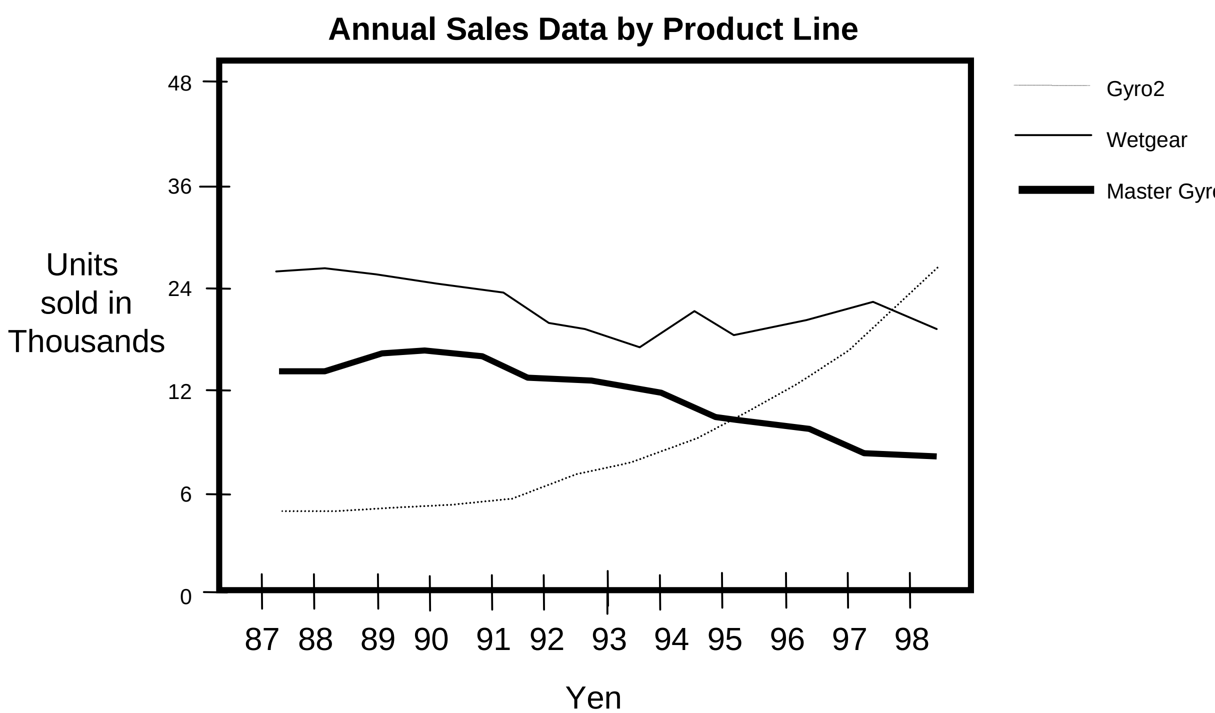
Guidelines:
- ✓ Conceptually; connect to the mid points of the various bars of a histogram
- ✓ If showing two groups on the same chart, use a dotted line for one and a solid line for the other

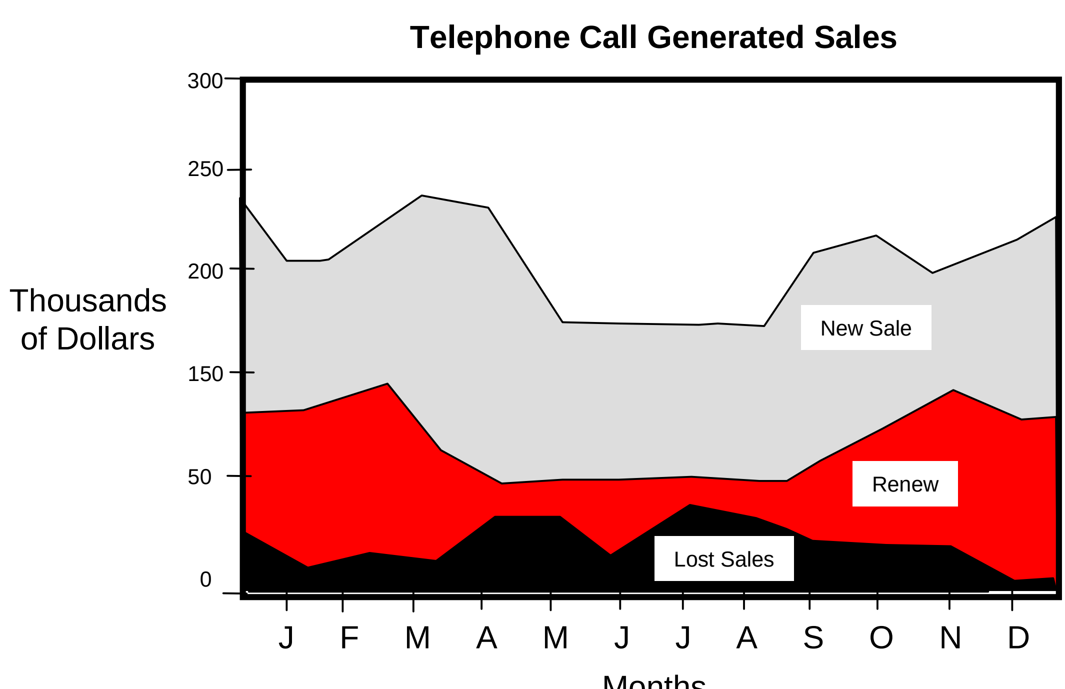
Cautions:
- ✓ When the first or last class value is a large number, beginning or ending the class category at zero distorts the appearance of the class value. In such cases, DO NOT use a frequency polygon.

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

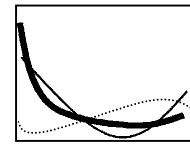
Figure 3.7: Example Line Graph and Area Chart





GUIDELINES

Constructing a Line Chart



Type of Data: ✓ Continuous (interval and ratio)

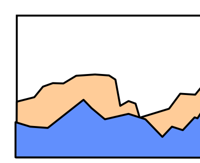
Uses:
- ✓ Emphasizes development patterns
- ✓ Provides clear picture for comparing changes over time

Guidelines:
- ✓ Time series values usually relate to periods (years, quarters, months, etc)
- ✓ For clarity, use 'ticks' to separate the periods
- ✓ Horizontal and vertical grid lines can be used to help with judging gradients and read levels
- ✓ The chart 'frame' should be large enough to provide space between the data and the frame
- ✓ Legend should not interfere with the graph
- ✓ Use symbols to mark lines sparingly
- ✓ Explain any abbreviations used in the chart

Cautions: ✓ A line chart becomes difficult to read if it has too many lines, particularly if they intersect

Figure 3.7 shows an example of a completed line graph and area chart.

Constructing An Area Chart



Type of Data: ✓ Continuous (interval and ratio)

Uses:
- ✓ Show development over time
- ✓ Shows the whole as well as its parts

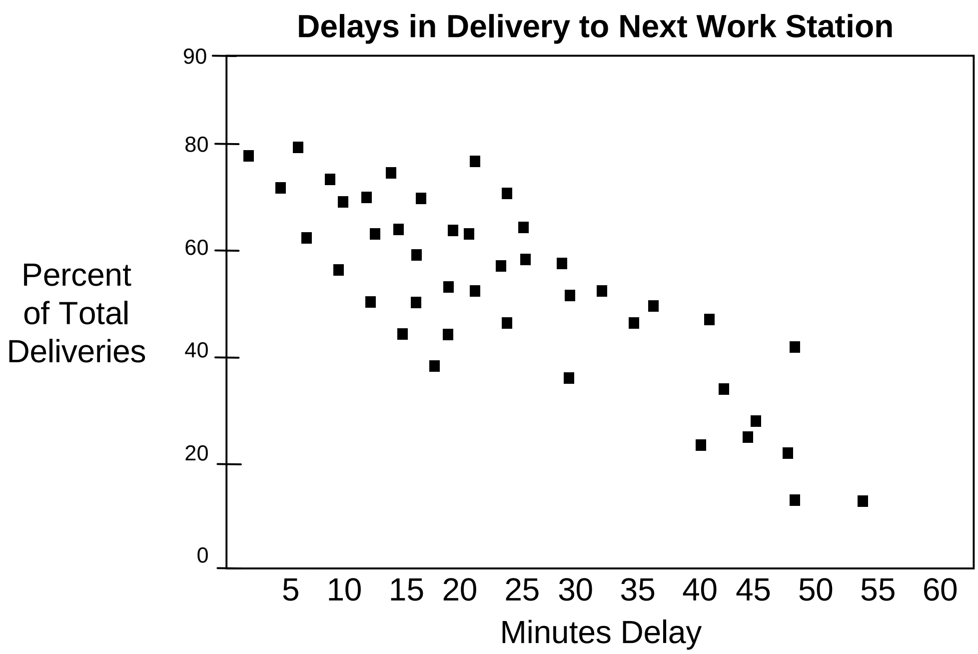
Guidelines:
- ✓ Use patterns or shading to distinguish the different areas
- ✓ Clearly state what each area represents
- ✓ If possible write legends directly in the areas
- ✓ Use brackets outside and to the right to improve clarity

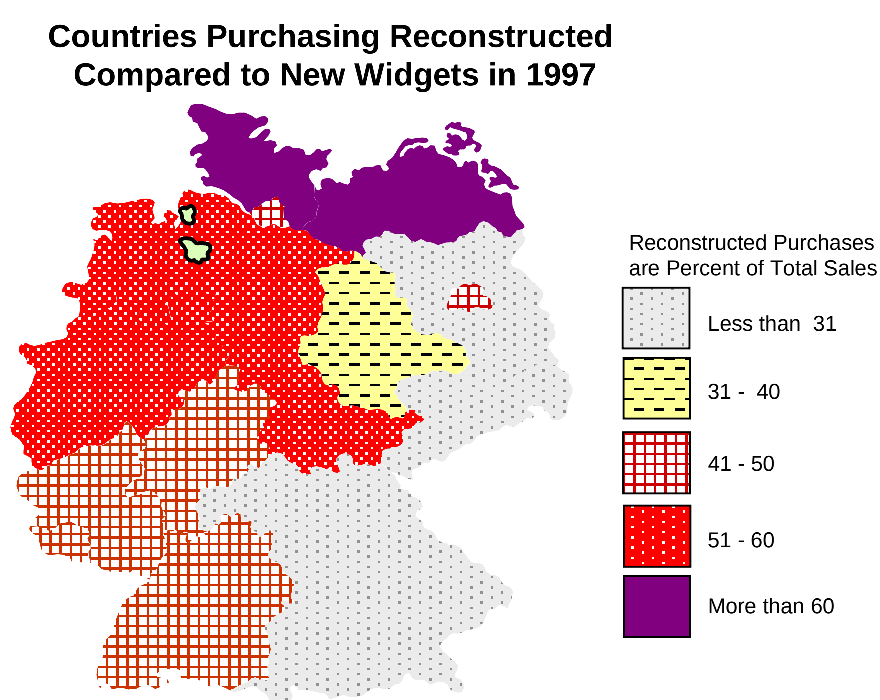
Cautions: ✓ In an area chart, it is easy to see value for the lower part. However, if the lower part varies greatly, it is difficult to determine the values of the other parts

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

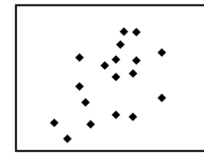
Figure 3.8: Example Scattergram and Choropleth Map





GUIDELINES

Constructing a Scattergram



Type of Data: ✓ Continuous (interval or ratio)

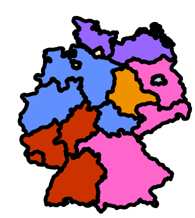
Uses:
- ✓ Shows how two variables co-vary
- ✓ Also called scatterplot or scatter diagram

Guidelines:
- ✓ Relationship is between two quantitative variables
- ✓ Scale the X & Y axes so the data fills the whole plot area
- ✓ Axes must be in proportion to variations in the variables
- ✓ Shows relationships, not levels, so it is not necessary to begin at zero

Cautions:
- ✓ When displaying more than one variable and there are many dots, it is better to draw several smaller charts

Figure 3.8 shows an example of a completed scattergram and choropleth map.

Constructing a Choropleth Map



Type of Data: ✓ Continuous (interval or ratio)

Uses:
- ✓ Good for showing ratios such as proportions, intensities, and averages for areas
- ✓ Ratio is divided into classes and represented by color, shading, or patterns
- ✓ Also known as “hatch maps” or shaded maps

Guidelines:
- ✓ Make sure shading and patterns are distinguishable
- ✓ Too much detail can provide a poor overview
- ✓ Don’t use borders around areas
- ✓ For small areas with a lot of detail, lift from main map and prepare an enlarged map with the relevant data

Cautions:
- ✓ Large areas can be over-represented -- they visually dominate the map
- ✓ Do not use for showing absolute numbers
- ✓ Great degree of subjective elements go into the production of such maps

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

Figure 3.9: The Parts of a Table

The parts that make up a table include:

Identifying Information	**Parts**
Table Number	Double Line
Table Title	Single Line
Box Head	Column Heads
Stubhead	Row Stubs
Source Note	Body (data)
Significant Notes	

Table number
Table title
Double line
Box head
Column heads
Stubheads
Row stubs
Body
Single lines
Source note
Significance note

Table 2.3

Number of Injuries in Last Ten Years

	YEAR											
Population	87	88	89	90	91	92	93	94*	95	96	97	Total
Men	33	29	16	18	18	17	16	13	9	8	5	182
Women	12	9	7	9	9	5	6	5	4	4	3	73
TOTAL	45	38	23	27	27	22	22	18	13	12	8	255

Source: OSHA Reports for 1987-1997

Note: *OSHA training plan implemented IN 1994

GUIDELINES

Constructing a Table

DEFINITION: A systematic array of words or numbers presented in rows and columns.

USED TO: Explain several data entries, reduce the amount of discussion needed in the text, and to show more clearly relationships of data.

RULES OF TABLE CONSTRUCTION

Order

- Order the rows and columns in a way that will make sense to the reader
- Put the largest row first or order data consistently - past, present, to future

Rounding Rules (Two digits)

- People do not understand more than two digits
- More than two digits are not justified statistically
- Accuracy beyond two digits is not usually necessary

Summaries

- It is important to summarize rows and columns to provide a standard for comparison

Figure 3.9 shows the parts of a table.

PARTS OF A TABLE

- **Table number** - for easy identification and referencing in the text
- **Title** - clear and brief description of the subject of the table
- **Body** - entries in the table
- **Stubhead** - names of rows
- **Column head** - names the column
- **Notes** - source notes, general notes, and specific notes

CONTENT OF TABLES:

- If items consist of words; capitalize only the first word, keep short (use symbols and abbreviations if possible), align to the left
- If items consist of numbers; align whole numbers to the right, align numbers with decimals points on the points
- If items consist of numbers and words; align the items on the left
- If items contain dollar amounts; align right or on the decimal point, use the dollar sign ($) only before the first amount and before the total amount
- If the figures in a column represent percentages; type a percent sign (%) directly after each figure unless the column heading clearly indicates the figures are percentages

USCG

WORKSHOP

SURVEY

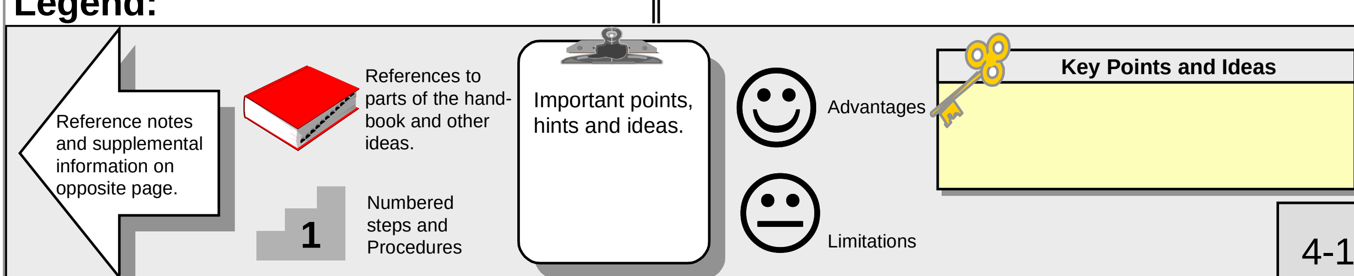
Section 4: Descriptive Statistics

CONTENTS

Overview

Guidelines

Legend:



NOTES AND SUPPLEMENTAL INFORMATION

OVERVIEW

Table 4.1: Descriptive Statistics Selection Guide

Characteristic of Interest	Unordered Qualitative Variable (nominal measurement)	Ordered Qualitative Variable (ordinal measurement)	Quantitative Variable (interval or ratio measurement)
Summary Measures (Organize)	Frequency Percent Cross-Tabs Pie Chart* Bar Chart*	Frequency Percent Cross Tabs Scattergram*	Frequency Percent Histogram* Frequency Polygon* Line Graph* Area Chart* Scattergram* Choropleth Map*
Central Tendency (Location)	Mode	Mode	Mode Median Mean Z-Score
Dispersion (Spread)	-----	-----	Range Interquartile Range Standard Deviation
Correlation (Association)	Cramer's Coefficient	Scattergrams Spearman's Coefficient	Scattergrams Pearson's Coefficient

*See Section 3-- Graphing Data

OVERVIEW

Introduction

Descriptive statistics summarize data.

A percentage is a descriptive statistic.

An average is a descriptive statistic.

A pie chart is used to present frequencies of data -- a descriptive statistic technique.

Descriptive statistics techniques are useful in helping a performance analyst to organize and summarize data. Descriptive statistics are particularly useful when there are large amounts of data that need to be interpreted.

For example, suppose a performance analyst administers a questionnaire to 75 materials handlers working in the transportation department of a large factory. The purpose of the questionnaire is to obtain information about accidents over the past 12 months. The answers on the questionnaire can be analyzed using descriptive statistics to answer questions such as: "What is the typical number of accidents for the workers?" Or, "How does this group's accident rate compare to national norms for materials' handlers in similar industrial settings?"

Descriptive statistics techniques that you are already familiar with are graphs, percentages and averages. Below, these and other techniques are discussed.

Using Descriptive Statistics

See Section 2 for further reading about 'population' and 'sample'.

Descriptive statistics are used to organize and summarize data from both populations and samples. Technically, statisticians use the term ***parameter*** to describe a statistical value from a population. When a value is based upon sample data, it is called a ***statistic***. Often, people call both values ***statistics***.

Table 4.1 is a selection guide for descriptive statistics covered in this section.

There is another convention that most statisticians follow to differentiate between descriptive statistics used for population data versus sample data. It has to do with ***symbols.*** For instance, the upper case "N" and lower case "*n*" are used to represent the number of cases in a set of data. For example, N = 100 for a population and *n* = 10 for a sample of the population.

A final concern in using descriptive statistics has to do with ***scales of measurement***. Performance analyst use a number of methods to collect data: document analysis, observations, focus group interviews, one-on-one interviews, questionnaires, etc. Different methods yield different types of data. The data can be classified according to scales of measurement -- the process of assigning numerals to the characteristics of people, things, or events using a set of rules. This in turn, impacts the type of statistical techniques that are used with the varying scales of measurement.

- ✓ ***Nominal*** measurement → naming or labeling data
- ✓ ***Ordinal*** measurement → rank order data
- ✓ ***Interval*** measurement → data with equal distance between numbers
- ✓ ***Ratio*** measurement → data with equal distance between numbers and an absolute zero

NOTES AND SUPPLEMENTAL INFORMATION

OVERVIEW

Figure 4.1: Important Statistical Symbols

SYMBOLS		MEANING
Population	Sample	
N	*n*	number of elements in a set
~~IIII~~ I	~~IIII~~ I	tally marks
f_x	f_x	frequency of variable “X”
P or %	p or %	percentage
3	3	summation sign
ΣX	ΣX	sum of values of a variable named “x”
X^2	X^2	square a value “x”
XY	xy	multiplication of value x and y
X_i	x_i	value of the i^{th} element in a set
$\overline{\sigma\quad}$	$\overline{\sigma\quad}$	square root
MO	*mo*	mode
MD	*md*	median
$\bar{X}$ or Φ	$\bar{x}$	arithmetic mean
R	R	range
Q	q	interquartile range
SD or Σ	s	standard deviation
P	r	coefficient of correlation
P^2	r^2	coefficient of determination
$Φ^2$	s^2	variance

OVERVIEW

Figure 4.1 lists the statistical symbols and their meanings as used in this section.

Computations

Most researchers and performance analysts recommend that users of statistics have some familiarity with the mathematical formulas and computations necessary to calculate basic statistical formulas. On the following pages you will find basic statistical formulas and illustrations using small sets of data. Also, reference the texts listed below in the section ***Further Readings***.

Access to computers has greatly reduced the burden on users of statistics. Computerized ***statistical software packages*** enable analysts to enter data, select the statistics to be computed, and get answers almost immediately without looking at the formulas and processes involved in the computations. The major statistical software packages which have versions for micro and lap-top computers are: IBM Scientific Subroutine Packages, Minitab, Statistical Analysis Systems (SAS), and Statistical Package for the Social Sciences (SPSS).

A cautionary note. The speed and ease of using statistical software packages makes possible for the basic level user to perform complex statistical procedures almost effortlessly. The concern is using procedures which are not fully understood by the needs analyst resulting in inappropriate analyses and incorrect interpretations of data. If a performance analyst does not understand the assumptions, conditions of use, and interpretation of the statistical procedures being used, ***DO NOT*** use the procedure. Or, seek consultation and technical assistance from a statistical consultant.

Further Readings

Kirk, R.E. (1990). Statistics: An introduction, (3rd ed.). Fort Worth, TX: Holt, Rinehart, and Winston.

Holcomb, Z. C. (1998). Fundamentals of descriptive statistics. Los Angeles, CA: Pyrczak.

Pyrczak, F. (1995). Making sense of statistics: A conceptual overview. Los Angeles, CA: Pyrczak.

Rosenfeld, R. (1992). The McGraw-Hill 36-hour business statistics course. New York, NY: McGraw-Hill.

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

Figure 4.2: Frequencies and Percentages

FREQUENCIES:

f_X = tally or count of X where X = variable (type of subject or case)

To what extent are you involved in identification of safety violations?	Tally Marks	f_x
A) Not involved	卌 I	$f_A = 6$
B) Conduct detailed safety audits	卌 卌 卌 卌 II	$f_B = 22$
C) Rely on managers' identification of safety problems	卌 卌	$f_C = 10$
D) Assist senior safety officers to conduct analysis	卌 卌 卌 II	$f_D = 17$
E) Other	III	$f_E = 3$
	Total	n = 58

PERCENTAGES:

$$P = \frac{f_X}{n} \times 100$$ where f_X = *frequency of a class*, n = *total number of scores*

	f_x	%
A	$f_A = 6$	$P = \frac{6}{58} \times 100 = 10.3\%$
B	$f_B = 22$	$P = \frac{22}{58} \times 100 = 37.9\%$
C	$f_C = 10$	$P = \frac{10}{58} \times 100 = 17.2\%$
D	$f_D = 17$	$P = \frac{17}{58} \times 100 = 29.3\%$
E	$f_E = 3$	$P = \frac{3}{58} \times 100 = 5.2\%$
	n = 58	

GUIDELINES

Frequencies, Percentages, and Cross-Tabs

When a performance analysis study produces large amounts of data, procedures are needed to summarize and depict the data so they are more readily comprehended. Techniques commonly used are frequency distributions, percentages and cross-tabs.

Frequency: f_X = tally or count of X where X = variable (subject or case)

Definition: The number of subjects or cases stated as absolute numbers. Its symbol is *f*. Note that the *f* is italicized.

Characteristics: It is most often used to describe the number of subjects or cases in subsets of a population or sample. For example, to describe how many workers in a company are hourly workers. Or, how many were absent last month due to work-related accidents.

Figure 4.2 shows how to calculate frequencies and percentages.

When to Use: With any set of raw data as a quick indicator of the subcategories of data. Often, frequencies are presented in data tables or graphs as the first descriptive statistics reported.

Percentage: P or % = $\frac{f_X}{n}$ x 100 where f_X = *frequency of a class*; n = *total number of scores*

Definition: Represents the number of subjects or cases per 100 that have a certain characteristic. Its symbol is P or %. Note that the P is <u>not</u> italicized.

Characteristics: Percentages convert frequency counts to a standard metric. Percentages are very useful in comparing one subgroup with another. For example, If we are told 12% of the workers in a company are managers, we know that for every 100 workers, 12 are managers. Similarly, if 67% are hourly workers, we know that for every 100 workers, 67 are hourly.

When to Use: Percentages are very useful in presenting categorical data. They are also helpful when comparing two or more groups of different sizes. In reporting percentages, it is useful to also report the underlying frequencies. Percentages can be mis-leading if not reported with the frequencies they represent.

Figure 4.3: Cross-Tab Analysis

Data with Two Categorical Variables

	Name	Gender	Position		Name	Gender	Position
1	Jackson	M	S	14	Rinker	F	S
2	Holcomb	M	H	15	Bertrand	M	S
3	Smith	F	S	16	Catterly	M	H
4	Ellis	M	S	17	Moenich	M	H
5	Wills	M	H	18	Adams	M	H
6	Goldman	M	H	19	Sayer	M	H
7	Abdel	M	H	20	Trout	M	S
8	Sanchez	M	H	21	Cappeletti	F	H
9	Pollard	M	H	22	Oshasha	F	H
10	Jordan	F	S	23	Sands	F	S
11	Pickering	F	S	24	Stone	F	S
12	Bond	F	S	25	Winfred	F	H
13	Stonge	F	S				

H = Hourly Position
S = Salary Position

Sort employees by making a mark in the appropriate cell.

		Position	
		Salary	Hourly
GENDER	F	~~////~~ /// 8	/// 3
	M	//// 4	~~////~~ ~~////~~ 10

2 X 2 Frequency Table: Cross-Tabulation of 25 Employees

GUIDELINES

Displaying Frequencies and Percentages:

See pages 22 and 23 in Section 3 on Graphing Data for guidelines on constructing a table.

Both frequency and percentage data are often depicted graphically. Pie charts, bar charts, histograms, and frequency polygons are the most commonly used graphics to depict frequencies and percentages. Guidelines for using these graphics are presented in ***Section 3*** titled **Graphing Data**.

Cross-Tab Analysis:

		WORKER TYPE	
		Salary	Hourly
Gender	M	9	87
	F	3	145

Figure 4.3 shows how to carry out a cross-tab analysis

Definition: Summarization in a table format of a set of categorical data consisting of two or more variables, with each variable having two or more categories. Also called **cross-tabulation** or **contingency table**.

Characteristics: Each cell in a table contains the number of responses in the appropriate cross-classification. The data are mutually exclusive -- that is each individual measure can go in only one cell. Association between variables is observed by examining the patterns of frequencies in the individual cells. It is sometimes helpful to note the frequencies in cells as percentages. The sum of the percentages in all of the cells would equal 100%.

Often such tables are used with ***dichotomous*** variables -- that is, variables measured using two nominal identifiers such as male/female, salary/hourly, technical training/academic training, etc. Dichotomies also arise when people are asked yes/no questions.

When to Use: For breaking data down into subgroups to analyze two or more variables at the same time. As an example, if we wanted to know how gender varied for salary versus hourly workers, we would create a ***2X2 cross tab table*** with gender (male/female) as the rows and worker type (salary/hourly) as the columns.

Figure 4.4: Calculating the Mode, Median and Mean

MODE: The most frequently occurring score

Scores:	7, 15, 7, 10, 5, 13, 12, 2, 11, 8, 5, 5
Place scores in order:	2, 5, 5, 5, 7, 7, 8, 10, 11, 12, 13, 15
Find the score that is repeated the most often.	2, $\boxed{5, 5, 5,}$ 7, 7, 8, 10, 11, 12, 13, 15
The mode:	**5** THE MODE

MEDIAN: The middle point in a distribution

Scores:	7, 15, 7, 10, 5, 13, 12, 2, 11, 8, 5, 5
Place scores in order:	2, 5, 5, 5, 7, 7, 8, 10, 11, 12, 13, 15
Find the middle score. If there are an even number of scores, then you must used the middle two scores to determine the median:	1 → 6 6 ← 1 2, 5, 5, 5, 7, $\boxed{7, 8,}$ 10, 11, 12, 13, 15
Compute the Median:	$\frac{7 + 8}{2} = \frac{15}{2} =$ **7.5** THE MEDIAN

MEAN: An Average

Scores:	7, 15, 7, 10, 5, 13, 12, 11, 2, 8, 5, 5
Find the sum of the scores:	7+15+7+10+5+13+12+11+2+8+5+5 = 100
Count the number of scores:	12
Compute the Mean by dividing the sum of scores by the number of scores:	100/12 = **8.33** THE MEAN

GUIDELINES

Measures of Central Tendency

It is often desirable to represent a set of data with a typical or representative number. Measures of central tendency condense a set of numbers into one number that typifies the middle of the set. Commonly used measures of central tendency are the mode, the median, and the mean.

Figure 4.4 shows how to calculate the mode, median, and mean.

Mode: Mo = highest occurring f_x where f_x = frequency of a value

Definition: The most frequently occurring individual value in a set of data.

Characteristics: Mode simply means “most popular.” When data are displayed in a bar graph, the bar with the highest frequency count will be the mode. Sometimes there are two “most popular” numbers. In such cases, the distribution of numbers is called bi-modal.

When to Use: The mode is most useful when the numerical values in a set are labels or categories -- that is, nominal scaled data.

Median: Md = middle value X where X = value of the middle observation

Definition: The middle point in a distribution of numbers when the data are arranged in numerical order from low to high.

Characteristics: The arranged numbers in a set are counted individually. If there is an odd number of numbers, then the “middle value” -- that is the number half-way from the low end and half-way from the high end of the listing is the median. If there is an even number of numbers, the median is half-way between the two middle values. The median splits a set of data into two equal-sized groups.

When to Use: It is best used with data sets with a few extremely high or extremely low scores compared to the rest of the scores.

Mean (Arithmetic Mean): $\bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ where $\sum_{i=1}^{n} X_i$ = sum of all the values; n = number of values

Definition: The sum of all of the values of the numbers in a set of data divided by the number of individual values.

Characteristics: It is the most frequently used measure of central tendency. It is commonly called the “average.”

When to Use: Best choice when data are more or less evenly distributed from lowest to highest values and there are no extreme values. Best used with interval or ratio scaled data.

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

Figure 4.5: How to Calculate Measures of Dispersion

Number of Accidents on Assembly Line
Over 20-Day Period

Day	Accidents	Day	Accidents
1	7	11	12
2	15	12	7
3	5	13	14
4	11	14	15
5	$\boxed{16}$	15	6
6	10	16	14
7	13	17	$\boxed{4}$
8	6	18	7
9	9	19	11
10	8	20	10
	100		100

CALCULATING THE RANGE

Method #1:

range = largest numerical value - smallest numerical value
= 16 - 4
= 12

Method #2:

Find the range using lower and upper limits. Although numerical scores often come in whole numbers, it is reasonable to think that these numbers are used for a range of possible scores. For example, a time of 57 seconds represents all times between 56.5 seconds and 57.5 seconds.

To find the upper and lower limits of any two numerical scores, subtract half the unit of measurement from the lowest score and add half the unit of measurement to the highest score. So, for the lowest score 4 accidents, the lower limit is 4.0 - .5 = 3.5. The upper limit score is 16 accidents, 16 + .5 = 16.5.

Range(R) = $X_u - X_L$ where X_u = upper limit and X_L = lower limit
= 16.5 - 3.5
= 13

CALCULATING THE INTERQUARTILE RANGE

Step #1: Place the values in order from low to high
Step #2: Divide the set of values into quarters or fourths
Step #3: For the values in the middle 50%, subtract the lower value (X_L) from the higher value (X_u)

Interquartile Range (Q) = 4 5 6 6 7 | 7 7 8 9 10 | 10 11 11 12 13 | 14 14 15 15 16

25% 50% 75%

= 13.5 - 6.5
= 7

GUIDELINES

Measures of Dispersion

Measures of dispersion are used to describe the amount of spread in a set of data. They describe how much a set of data varies. Measures of dispersion are also useful when comparing one set of data with another in terms of their variability. The three most frequently used measures of dispersion are the range, the interquartile range, and the standard deviation.

RANGE: $R = X_u - X_L$ where X_u = upper limit number, X_L = lower limit number

Definition: Difference between the largest and smallest value in a set of data.

Figure 4.5 shows how to calculate the range and the interquartile range.

Characteristics: Based upon the two most extreme scores. If the distribution of scores is skewed (i.e., has large outlying scores), it is a weak measure of dispersion.

When to Use: As a quick indicator to see how spread the data are or reported in conjunction with other measures of dispersion.

INTERQUARTILE RANGE: Q = $X_u - X_L$ of middle 50% of scores

Definition: The range of the middle 50% of scores or ranked data. If all scores are put in order from low to high and the lower and top 25% are ignored, the range of the remaining scores is the interquartile range.

Characteristics: It is not as sensitive as the range to extreme values in a data set.

When to Use: When data in a set have extreme values, use the interquartile range. Also, it is customary to use the interquartile range as the measure of dispersion when the median is used as the measure of central tendency.

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

Figure 4.6: Calculating the Standard Deviation

$$SD = \sqrt{\frac{\Sigma x^2}{N}}$$

Step #1: Count the number of cases = N

Step #2: Sum the value of the cases = 3 X

Step #3: Calculate the arithmetic mean = $\bar{X}$

Step #4: Find the deviation scores (mean - raw score) = $\bar{X}$ -X

Step #5: Square each deviation score = X^2

Step #6: Sum the squared values (called sum of squares) = Σx^2

Step #7: Divide the sum of squares by N, the number of raw scores

Step #8: Calculate the square root of the value

Accidents (X)	Deviations $(\bar{X} - X) = X$	Deviation Squared (X^2)
16	6	36
15	5	25
15	5	25
14	4	16
14	4	16
13	3	9
12	2	4
11	1	1
11	1	1
10	0	0
10	0	0
9	-1	1
8	-2	4
7	-3	9
7	-3	9
7	-3	9
6	-4	16
6	-4	16
5	-5	25
4	-6	36

N = 20

ΣX = 200

ΣX^2 = 258

Arithmetic mean = $\bar{X} = \frac{\Sigma X}{N} = \frac{200}{20} = 10$

$$SD = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{258}{20}} = \sqrt{12.9} = 3.59$$

GUIDELINES

STANDARD DEVIATION:

$$SD = \sqrt{\frac{\Sigma X^2}{N}}$$

where

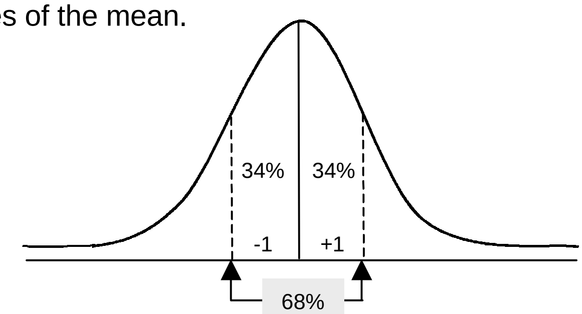
ΣX^2 = sum of squares of individual deviation from mean

N = number of items

Figure 4.6 shows how to calculate the standard deviation

Definition: The average of the deviations of a set of scores from their mean score. It measures how much scores differ from the mean score of their group.

Characteristics: It is the most popular measure of dispersion. The more spread out scores are from their mean score, the larger the standard deviation. It has special meaning when considered in relation to the normal curve. Generally, if a set of scores is normally distributed, about two thirds of the scores will be within one standard deviation on both sides of the mean.



When to Use: When the mean is reported, it is common to also report the standard deviation.

Correlation

Correlation deals with the extent to which two variables are related across a group of people, objects, or events. A ***variable*** is a characteristic or trait on which people, objects or events differ or vary.

For example, workers vary in the amount of formal education they have completed. Also, people vary in how much money they earn working. Are formal education and money earned related? Often, yes. Workers with more formal education earn more money working. Therefore, we can say that there is a direct positive relationship between the two variables.

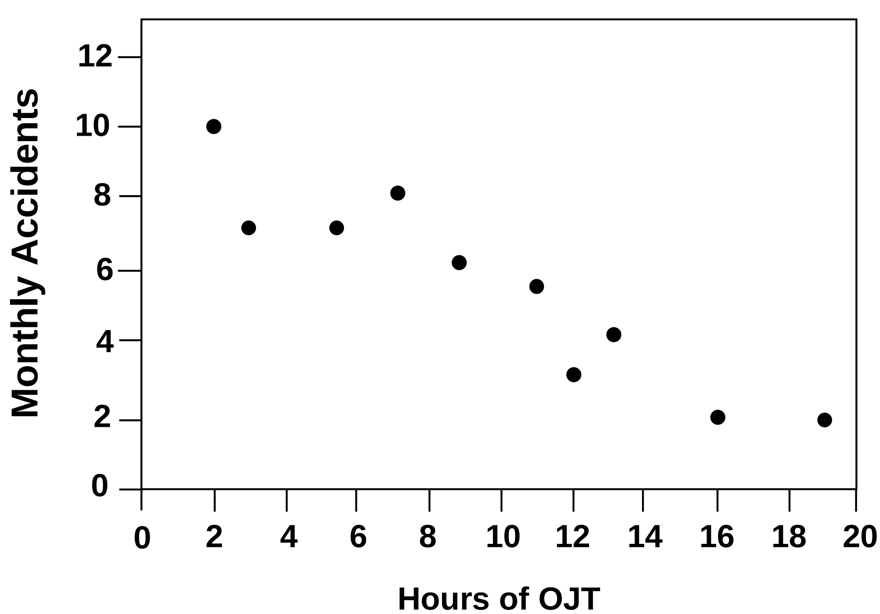
However, we know there are exceptions. Some people with very little formal education make more money working than some people with much formal education. All of us have heard stories of the high-school drop out who earns $200,000 per year in his/her construction business and the high school teacher with a masters degree earning $25,000 per year. The number and size of the exceptions are important in describing the relationships between two variables. Below are two methods for describing relationships between two variables -- ***scattergrams*** and the ***Pearson product-moment correlation coefficient***.

Figure 4.7: Creating a Scattergram

Hours of On-The-Job Training and Monthly Accidents

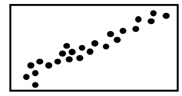
	Hours of OJT *X* variable	Monthly Accidents *Y* variable
Employee A	16	2
Employee B	5	7
Employee C	9	6
Employee D	2	10
Employee E	12	3
Employee F	19	2
Employee G	11	5
Employee H	7	8
Employee I	13	4
Employee J	3	7

Hours OJT and Monthly Accidents



GUIDELINES

Scattergrams:

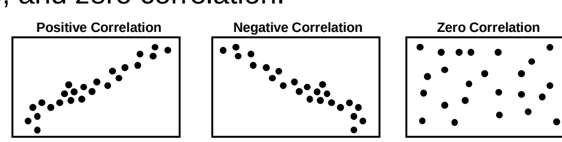


Definition: A graphical depiction of the relationships of two variables. Also called ***scatterplot*** or ***scatter diagram.***

Figure 4.7 shows how to construct a scattergram.

Characteristics: It yields a picture on a two-dimensional grid that helps the needs analyst see how two variables relate to each other. In looking at a scattergram, one of three general patterns will emerge. If the pattern of marks (e.g., dots) on the grid are along an imaginary line that slopes upward from left to right, there is a ***positive correlation***. In such cases, high values or scores on one variable tend to be high on the other variable and low values on one variable tend also to be low on the other. If the pattern of marks (e.g., dots) on the grid are along an imaginary line that slopes downward from left to right, there is a ***negative correlation***. In such cases, high values or scores on one variable tend to be low on the other variable and low values on one variable tend to be high on the other. If the pattern of marks (e.g., dots) is random on the grid, there is ***zero correlation***. In such cases, there is no regular pattern in the relationships between the values of the variables. The following scattergrams depict positive, negative, and zero correlation.

See pages 20 and 21 in Section 3 on Graphing Data for guidelines on constructing scattergrams.



When to Use: Constructing and looking at a scattergram is the first step in correlation analysis. It gives you an indication if there is an apparent relationship between two variables.

A Word of Caution! Establishing a correlation between two variables **DOES NOT** mean that a ***causal relationship*** has been established. In short, merely observing or even suspecting that one variable is positively or negatively correlated with another does not mean that causation is warranted. Consider the hypothetical that workers who are more receptive to taking orders from supervisors get more paid overtime work. Although a positive relationship was found, there could be several causal explanations. For instance, those who get overtime may know the supervisors on a personal basis, that is, be their friends. Or, overtime may be on the basis of seniority. Workers with more seniority may be older employers, and part of a work culture that is conditioned to “taking orders from supervisors.” Other explanations are plausible.

In order to determine a ***cause-and-effect relationship*** between associated variables, a controlled experiment is needed in which variables are treated differently to see what relationships emerge.

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

Figure 4.8: Calculating the Pearson Product-Moment Correlation Coefficient

Employee	Hours OJT *X*	Accidents *Y*	X^2	Y^2	X*Y
A	16	2	256	4	32
B	5	7	25	49	35
C	9	6	81	36	54
D	2	10	4	100	20
E	12	3	144	9	36
F	19	2	361	4	38
G	11	5	121	25	55
H	7	8	49	64	56
I	13	4	169	16	52
J	3	7	9	49	21
N=10	$\Sigma X=97$	$\Sigma Y=54$	$\Sigma X^2=1219$	$\Sigma Y^2=356$	$\Sigma XY=399$

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}$$

$$r = \frac{(10)(399) - (97)(54)}{\sqrt{[(10)(1{,}219) - (97)^2][(10)(356) - (54)^2]}}$$

$$r = \frac{3990 - 5238}{\sqrt{[12190 - 9409][3560 - 2916]}} = \frac{-1248}{\sqrt{[2781][644]}}$$

$$r = \frac{-1248}{\sqrt{1790964}} = \frac{-1248}{1338} = -.93$$

WHAT CAN BE CONCLUDED

This is a very strong negative correlation. It should be noted that most correlations will not be this strong. Based on the negative correlation coefficient, we can say that in most cases, the more hours an employee spends receiving on-the-job training, the fewer accidents the employee will have per month.

It is important to note that just because we have established a strong correlation, it does not necessarily mean that a **causal relationship** has been established. In order to determine cause-and-effect, an experiment is always needed.

Pearson Product Moment Correlation Coefficient:

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}$$

where

Σx =	sum x scores	Σx^2 =	sum (x)(x) scores
Σy =	sum y scores	Σy^2 =	sum (y)(y) scores
Σxy =	sum (x)(y)scores	N =	number of paired scores

Definition: Indicates the degree of linear association of two numeric variables. Represented by the letter *r*.

Characteristics: The correlation coefficient has a possible range from -1 to +1. The sign of the*r* represents the direction of the association. Positive (+) for a direct association and negative (-) for an inverse association. A value of +1 or -1 indicates a perfect linear relationship between two variables. A + or - correlation close to zero (0) indicates littler or no correlation. It is a widely used statistic and often is cited in needs analysis studies.

Figure 4.8 shows how to calculate the Pearson product-moment correlation coefficient.

When to Use: It is useful for summarizing the direction and strength of the association between two variables. The **Pearson *r*** is used for variables measured using a continuous scale -- interval or ratio data. The relationship between the two variables needs to be roughly linear, that is follow a straight line. If the data yield a pattern that overtly curves, it is NOT a good statistic to use. It is best to produce a scattergram to visually detect whether or not the data are more linear or curvilinear (that is, follow a straight line pattern or a curved line pattern).

Interpreting the Pearson *r:* The Pearson *r* is NOT a proportion. It cannot be multiplied by 100 to obtain a percentage. A Pearson $r = .60$ does not convert to 60%. To convert correlation to percentages, it is necessary to convert a Pearson *r* to the ***coefficient of determination***. Its symbol is r^2 and is computed simply by squaring the value of *r*. Therefore, the coefficient of determination for $r = .60$ is $r^2 = (.60)^2 = .36$ or 36%. Simply stated, the coefficient of determination tells us how much of the variance between two variable is accounted for. For $r = .60$, 36% of the variance between two variables (that is, correlation) is accounted for. The Figure on the next page shows corresponding coefficients of determination for various values of Pearson *r*.

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

Figure 4.9: Approximate Interpretation of Correlation Coefficient Ranges

Correction Value			Approximate Interpretation
- 1.0			**Perfect negative correlation**
-.8	to	-1.0	Very high degree of negative correlation
-.6	**to**	**-.8**	**High degree of negative correlation**
-.4	to	-.6	Medium degree of negative correlation
-.2	**to**	**-.4**	**Low degree of negative correlation**
+.2	to	-.2	Probably no correlation
+.2	**to**	**+.4**	**Low degree of positive correlation**
+.4	to	+.6	Medium degree of positive correlation
+.6	**to**	**+.8**	**High degree of positive correlation**
+.8	to	+1.0	Very high degree of positive correlation
+1.0			**Perfect positive correlation**

GUIDELINES

Figure 4.9 provides guidelines for interpreting a range of correlation coefficient.

Alternative Formulas:

For rank level data (ordinal scaled measures) the recommended correlation coefficient is the ***Spearman Rank Correlation Coefficient***. For dichotomous or qualitative data (nominal scaled measures) use ***Cramer's Measure of Association***. These formulas and examples of their applications can be found in most basic statistics books as well as statistical software packages.

Spearman's r_s:

$$r_s = 1 - \frac{6 \sum D^2}{N^2 - N}$$

where

D = difference between a subject's rank on two variables.

N = number of subjects

Cramer's $\hat{V}$:

$$\hat{V} = \sqrt{\frac{X^2}{n(s-1)}}$$

where

X^2 = Chi-square

n = number of observations

s = smaller of the number of rows or columns

A Final Word on Descriptive Statistics

Descriptive statistics is the branch of statistics that involves summarizing, tabulating, organizing, and graphing data for the purpose of describing a population or sample of people, objects, or events that have been observed or measured. In choosing the proper descriptive statistic, a performance analyst must consider the nature of the measures taken and the purpose for taking the measures. When in doubt, it is best to reference a basic textbook on statistics and to seek the advice of an expert.

USCG

WORKSHOP
U
R
V
E
Y

Section 5: Inferential Statistics

CONTENTS

Overview

Guidelines

Legend:

Reference notes and supplemental information on opposite page.

References to parts of the hand-book and other ideas.

1 Numbered steps and Procedures

Important points, hints and ideas.

Advantages

Limitations

Key Points and Ideas

NOTES AND SUPPLEMENTAL INFORMATION

OVERVIEW

Table 5.1: Inferential Statistics Selection Guide*

Number of Samples	Unordered Qualitative Variable (Nominal Measurement)	Ordered Qualitative Variable (Ordinal Measurement)	Quantitative Variable (Interval or Ratio Measurement)
One-Sample	➤One-sample Z-test for a population proportion ➤Chi-square goodness of fit test	➤ One-sample Z-test for a population proportion ➤ Chi-square goodness of fit test	➤ One-sample t-test for a mean ➤ One-sample t- and Z-tests for a population correlation
Two Independent Samples**	➤Chi-square test	➤ Mann-Whitney *U* test ➤ Kolmogorov-Smirnov two-sample test	➤ t-test for independent samples
Two Dependent Samples***	➤McNemar test for the significance of changes	➤ Wilcoxon matched pairs signed-ranks test	➤ t-test for correlated samples
Multiple Independent Samples**	➤Chi-square test for K independent samples	➤ Kruskal-Wallis one-way analysis of variance	➤ Analysis of variance
Multiple Dependent Samples***	➤Cochran Q test	➤ Friedman two-way analysis of variance	➤ Repeated measures analysis of variance

* Once the number of samples and level of measurement are determined, choose the proper inferential statistical test from the table. Use one of the textbooks listed in “Further Readings” for a full explanation of the test.

** ***Independent Sample*** = selection of the elements in one sample is not affected by the selection of elements in the other. Example: randomly choose from two populations or use random selection procedure to assign elements from one population to two samples.

*** ***Dependent Sample*** = selection of the elements in one sample is affected by the selection of elements in the other. Example: repeated measures using the same subjects, subject matching, selecting pairs of business partners, etc.

OVERVIEW

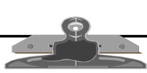
Introduction

Inferential statistics are useful in helping an analyst to generalize the results of data collected from a sample to its population. As the name implies, the analyst uses ***inferential statistics*** to draw conclusions, to make ***inferences***, or to make predictions about a given population from sample data from the population.

Table 5.1 is a selection guide for inferential statistics.

For example, suppose a performance analyst administers a questionnaire to 100 randomly chosen hourly employees from a population of 550 hourly employees. The purpose of the questionnaire is to obtain information about preferences for work scheduling. Seventy-three of the 100 surveyed employees (73 percent) said they would prefer working 4 days for 10 hours over 5 days for 8 hours. When the needs analyst reports that about 73% of all hourly employees would prefer working a 4-day, 10-hour shift, the analyst is making an ***inference*** from the sample. In short, the analyst is making an ***estimate*** of the population parameter. The sample percentage is an estimate of the true parameter percentage.

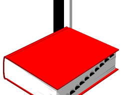
Using the above example -- let's call it data from plant A. An analyst might be interested in comparing the plant A percentage with workers in plant B. In plant B, 100 randomly chosen hourly employees were administered the same questionnaire. The data indicate that 68 of the 100 surveyed employees would prefer working 4 days for 10 hours. There are 600 total hourly employees in plant B. As in the example above, the 68% is an ***estimate*** of the population parameter. Comparing the percentages from the two plants, involves a form of ***hypothesis testing***.



Parameters come from ***P***opulations.

Statistics come from ***S***amples.

Estimating population parameters and ***hypothesis testing*** are two basic ways that inferential statistics are used. Any type of generalization or prediction about one or more groups (***populations***) based upon parts of the groups (***samples***) is called an ***inference***. Remember, the groups can be people, objects, or events.



See page 3 in Section 4 for further information about varying types of data.

Inferential statistics techniques are not as familiar as descriptive techniques to most people. The most common techniques used in needs and performance analysis studies are t-tests, analysis of variance (ANOVA), and Chi-Square. As with descriptive statistics, there are many types of inferential statistics tests. The variety is necessary to accommodate varying types of data (that is, nominal, ordinal, interval and ratio scaled data), the size of samples, and the number and relationships of the samples being analyzed.

Relationships of Descriptive and Inferential Statistics

To begin with, ALL population statistics are descriptive. Technically, these statistics are called ***parameters*** because they describe certain characteristics of a ***population***. They summarize and describe relationships between two or more variables.

NOTES AND SUPPLEMENTAL INFORMATION

OVERVIEW

Figure 5.1: Relationship of Descriptive Statistics to Inferential Statistics

Population →	**Sample** →	**Descriptive Statistics** →	**Inferential Statistics**
5,000 employees need to be tested using a performance test	Analyst randomly samples 500 employees and gives them the test	Used to summarize the sample scores from the 500 employees	Based on descriptive statistics use inferential statistics to estimate summary scores for the entire population of 5,000 employees

Figure 5.2: The Logic of Hypothesis Testing

1) The XYZ company has an extensive and successful training program. The training director is asked to provide training support for a new plant being built two miles down the road.

2) One year into providing training for the new plant, training personnel complain that the workers in the new plant are bored with training and always seem ahead of the instruction, no matter the subject matter or method of delivering the instruction.

3) After much searching for an explanation, the training personnel reach a consensus on the following: "It is just seems that we are dealing with a much more highly educated group of workers in the new plant."

4) This statement is translated into the following statement of conjecture: "We bet the group of workers in the new plant has more years of formal education than the group in the old plant." The training director says: "Let's test the idea."

5) The data gathering budget is limited, so it is decided that only 500 of the 5,000 employees in the new plant and 500 of the 5,850 employees in the old plant can be interviewed about their formal education. What is really being called for is a **test of a hypothesis**.

6) The training group begins with the following statement: "The average number of years of formal education of the employees in the new plant is the same as the average number of years of formal education of employees in the old plant. This type of statement of belief is known in statistical language as a **null hypothesis**.

7) It is a statement or belief about the value of **population parameters**. In this case, the average number of years of formal education of the workers in the new and old plants.

Continued on Page 5 - 6

OVERVIEW

Figure 5.1 shows the relationships between descriptive and inferential statistics

Similarly, when statistics are used to summarize and describe data from one or more samples, they too are called descriptive. Technically, these statistics are called ***descriptive statistics*** because they describe certain characteristics of ***samples***. They are NOT used to make generalizations, they simply summarize the values from which they are generated. The difference between parameters and descriptive statistics is that descriptive statistics are calculated using values from samples and parameters are calculated from values from populations. Both are used to ***describe*** a set of data.

When one wishes to generalize findings from a sample to a population, inferential statistics are used. By using samples that are just fractions or parts of one or more populations, an analyst can make timely and relatively inexpensive generalizations about large numbers of people, objects, or events. In doing so, the analyst is ***estimating.*** Comparing two or more estimates from different populations or from sub-parts of a population involves hypothesis testing. Understanding ***hypothesis testing*** is central to using inferential statistics.

The Logic of Hypothesis Testing

Often, analysts are searching for differences between people, objects, or events and for explanations for the differences they find. This is often done by ***testing hypotheses***. A hypothesis is a statement about one or more population parameters. Such statements are either true or false. The purpose of hypothesis testing is to aid decision makers, including performance analysts, to decide to accept or reject statistical findings from data obtained from sample data.

Figure 5.2 walks you through the logic of a hypothesis test.

Hypothesis testing is necessary because statistical values obtained from a sample may be close to the actual population value, but they will not be exactly the same. This error is called ***sampling error***. Technically, sampling error refers to the inherent variation between a score or value describing a characteristic derived from a sample and the actual score or value of the characteristic in the population from which the sample was drawn. When the error is due to chance or luck, the laws of probability can be used to assess the possible magnitude of the error.

Returning to the example above, the sample data indicated that 73% of the hourly workers in plant A and 68% of the hourly workers in plant B preferred working 4 days for 10 hours. This suggests that a higher percentage of workers in plant A prefer the 4 day 10 hour option -- 73% to 68%. Given that both percentages were derived from samples, it is possible that the difference obtained is due only to errors resulting from random sampling. In other words, it is possible that the population percentage for plant A workers and the population percentage for plant B workers are identical and the difference we found in sampling the two groups of workers is due to chance alone.

NOTES AND SUPPLEMENTAL INFORMATION

Figure 5.2: The Logic of Hypothesis Testing (Continued)

8) In hypothesis testing, belief in the validity of the null hypothesis continues unless evidence collected from one or more samples is strong enough to make the continued belief unreasonable.

9) If the null hypothesis is found to be unreasonable, we deem it false. Then, we accept that an **alternative hypothesis** must be true.

10) In this case, the alternative hypothesis is that the average number of years of formal education of the workers in the new plant is greater than the average number of years of formal education of the workers in the old plant. Remember that the training group started out believing the workers in the new plant have more formal education than those in the old plant. However, we made the converse assertion -- the formal education level of the workers in the new and old plants would be the same.

11) We attempt to **"nullify"** this assertion. In so doing, "the law of the excluded middle" is applied. This means that a statement must be either **true or false.** We do this because one cannot directly prove that the formal education levels of the two groups is different. We cannot get at the question directly, but we can indirectly.

12) The null hypothesis is stated in an absolute form -- **"no difference."** We can prove or disprove an absolute. If disproved, the alternative is then plausible.

13) In our samples, we find the formal education level of the employees in the new plant is greater than that of the employees in the old plant. The null hypothesis test will tell us whether the difference is due to random variability due to the samples selected or whether there is a probability that the difference is real.

14) If it is real, with a degree of probability that we are satisfied, we accept the **alternative hypotheses** and conclude that the average number of years of formal education of the workers in the new plant is greater than the average number of years of formal education of the workers in the old plant.

This possibility is known as the ***null hypothesis***. For differences between two percentages, it says:

> ***The true difference between the percentages (in the population) is zero.***

Symbolically, the statement is expressed as follows:

H_o: $P_1 - P_2 = 0$

where
- H_o is the symbol for the null hypothesis
- P_1 is the symbol for the population percentage for one group
- P_2 is the symbol for the population percentage for the other group

Figure 5.2 continues to walk you through the logic of hypothesis testing.

Other ways to state the null hypothesis are:

> ***There is no true difference between the percentages.***

The null hypothesis can also be stated in a positive form:

> ***The observed difference between the percentages was created by sampling error.***

As stated above, often the analyst is searching for differences among people, objects, or events. Therefore, most performance analysis studies are NOT undertaken to confirm the null hypothesis. However, once samples are used in a study, the null hypothesis becomes a possible explanation for any observed differences.

Sampling error is due to the inherent variation between an estimate of some characteristic computed from a sample and the actual value of the characteristic in the population from which the sample was drawn.

Performance analysts and decision makers will often have their own professional judgment or hypothesis. It is usually inconsistent with the null hypothesis. It is called a ***research hypothesis***.

Returning to the preferences for work scheduling example above, a performance analyst may believe that there will be differences in the responses of hourly workers in plants A and B on this matter, However, the analyst is not willing to speculate in advance which group of workers will be more or less in favor of a change. In research terminology, this is called a ***nondirectional research hypothesis***.

Symbolically, the statement is expressed as follows:

H_1: $P_1 \neq P_2$

where
- H_1 is the symbol for an alternative hypothesis which in this case is a nondirectional research hypothesis
- P_1 is the symbol for the population percentage for one group
- P_2 is the symbol for the population percentage for the other group

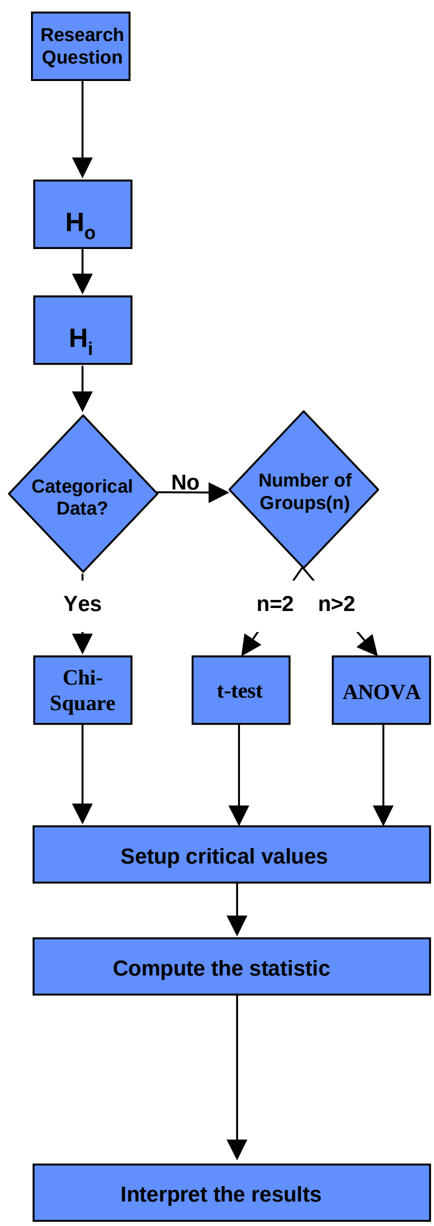
NOTES AND SUPPLEMENTAL INFORMATION

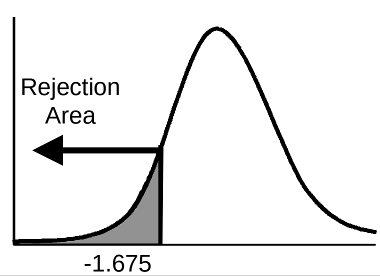
Figure 5.3: Steps of Hypothesis Testing

OVERVIEW

Prior to a new safety program, the average number of on-the-job accidents per day from 30 randomly chosen days over a year was 4.5 with a standard deviation of 1.9 days. To ascertain if the safety program had been effective, the number of accidents was recorded for 30 randomly chosen days over a one-year period after implementation of the safety program. The average number of accidents per day = 3.7 with a standard deviation = 1.3 days.

FLOWCHART



STEPS	EXAMPLE
STEP 1	As an analyst, you want to know if the new safety program is effective. After the safety program, is the average number of accidents per day lower than before the implementation of the program?
STEP 2	Null hypothesis: the mean numbers of accidents before and after the program are the same. Alternative hypothesis: the mean number of accidents before the program is greater than the mean number of accidents after the program. This is a directional hypothesis.
STEP 3	•Characteristic of data = independent samples &ratio data •Number of groups = 2 samples = 30(days each), df = 58 •Selection of test = t test •Directional hypothesis (H_1: $U_1 > U_2$) •Level of significance at .05 •Critical value = -1.675 (from a table for one-tailed t-test)  If observed value < -1.675, null hypothesis is rejected.
STEP 4	$t\ value = \frac{\text{mean before - mean after}}{\text{standard error of the difference between means}}$ $t = \frac{(3.7 - 4.5)}{.4203} = -1.90$ Value of t test statistic = -1.90 (less than critical value of -1.675). Therefore, reject the null hypothesis.
STEP 5	The difference is statistically significant at the .05 level(t=-1.90, df = 58). The safety program makes a difference, but is it practically significant? The analyst will need to ascertain the cost-effectiveness or other important attributes of the program.

OVERVIEW

On the other hand, what if the analyst knows in advance that plant B has more female hourly workers with school-age children than plant A? And further, these workers have previously expressed a desire to be home with their children after school, and working eight hours per day accommodates this need. The analyst may want to hypothesize that fewer hourly workers in plant B will want to change to a 4-day, 10-hour shift working day. In this case, the analyst would state a ***directional hypothesis***.

$\mathbf{H_1: P_1 > P_2}$ where H_1 is the symbol for an alternative hypothesis which in this case is a directional research hypothesis

P_1 is the symbol for the population percentage for the group hypothesized to have a higher percentage (in this case the hourly workers from plant A)

P_2 is the symbol for the population percentage for the other group (in this case, the hourly workers from plant B)

Figure 5.3 illustrates the steps of hypothesis testing.

Making Decisions About the Null Hypothesis

Inferential statistical tests about null hypotheses are designed to provide a ***probability that the null hypothesis is true***. The symbol is a lower-case italicized letter ***p***. For instance, if we find that the probability of a null hypothesis in a given situation is less than 5 in 100, this would be expressed as $\boldsymbol{p < .05}$. This means that it is quite unlikely that the null hypothesis is true. We are not certain, but the chances of the null hypothesis being true are less than 5 out of 100. In needs analysis studies, when the null hypothesis is less than 5 in 100, it is conventional to ***reject the null hypotheses***.

Another way that social science researchers express the rejection of the null hypothesis is to declare the result to be ***statistically significant***. In reports, the analyst would say:

The difference between the percentages is statistically significant.

Or, if mean scores are being compared, the word “percentages” is replaced by the word “means” to say:

The difference between the means is statistically significant.

These statements indicate that the analyst has rejected the null hypothesis. In needs analysis studies, frequently the ***p* value** of less than .05 is used. Other common *p* values are:

$\boldsymbol{p < .01}$ (less than 1 in 100)

$\boldsymbol{p < .001}$ (less than 1 in 1,000)

NOTES AND SUPPLEMENTAL INFORMATION

Figure 5.4: Meanings of Significance

The reporting of statistical significance is concerned with whether a result could have occurred by chance. If there is a probability that a result is statistically significant, then the analysts must decide if it is of practical or substantive significance. Practical significance has to do with whether or not a result is useful.

Statistical significance means that we have a value or measure of a variable that is either larger or smaller than would be expected by chance.

A large sample size often leads to results that are statistically significant, even though they might be inconsequential.

Statistical significance does not necessarily imply substantive or practical significance.

The meaningfulness of the attained level of significance, given that it must first be statistically significant, is best determined by key audiences, interest groups, and those directly impacted by a study.

In applied research situations, such as performance or needs analysis, practical significance means that there are effects of sufficient size to be meaningful.

There are no objective means to set a practical or substantive level of significance. Whether the level of outcome attained is important or trivial is arbitrary.

Before changes are made in programs, policy, etc., based on statistically significant data, acceptable practical significance levels <u>must</u> be pre-stated.

The factors that are important in establishing practical significance can be many and varied. Examples include:

- Cost factors or dollar implications
- Issues of disruptions caused by change
- Social and political factors
- Importance as viewed by interested groups

OVERVIEW

The Meanings of Significance

In the context of statistics, the word ***significance*** by itself refers to the degree to which a research finding is meaningful or important. The words ***statistical significance***, when used together, refer to the value or measure of a variable when it is larger or smaller than would be expected by chance alone. Statistical significance comes into play when the needs analyst is making ***inferences*** about population parameters from sample statistics.

There is a concern that overrides statistical significance. It has to do with ***practical*** or ***substantive significance***. This occurs when a research finding reveals something meaningful about what is being studied. In short, it presents a ***"so what"*** question. A finding may be statistically significant, but what are the substantive or program implications of the finding to the needs analysis study?

Figure 5.4 provides explanations about substantive and practical significance.

For example, suppose an analyst took a large sample of hourly workers from a plant in California and one in Indiana. In comparing the data, the analyst finds that the average age of the California workers is 33 years and the Indiana workers is 36 years. If the samples were large and representative, this 3-year age difference would be unlikely to be due to chance or sampling error alone. It would be ***statistically significant***. However, it would be hard to find a reason that a 3-year difference in age would have any ***practical*** or ***substantive significance (meaning)*** about hourly workers in the two states.

It is important to note that if a statistical finding is not statistically significant it cannot be substantively significant. Also, it is important in needs analysis studies to state prior to data collection and analysis the level of significance that must be reached to be of practical or substantive significance. In the case of the California and Indiana workers, a 5 or 10 year age difference may be meaningful if the older workers are nearing the age of retirement.

Computations

Most analysts will need some level of expert advice in carrying out studies involving inferential statistics. Reference the texts below and use the power and speed of a statistical software package to input, process, and produce data. If an inferential statistical procedure used is not fully understood, ***DO NOT*** use the procedure.

Further Readings

Kirk, R.E. (1990). Statistics: An introduction, (3rd Edition). Fort Worth, TX: Holt, Rinehart, and Winston.

Hinton, P. R. (1996). Statistics explained: A guide for social science students. New York, NY: Routledge.

Pyrczak, F. (1996). Success at statistics: A worktext with humor. Los Angeles, CA: Pyrczak.

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

Figure 5.5: Confidence Interval: An Example

Suppose you work in Human Resources at a company. You plan to survey employees to find their average days absent from work per year. By random sampling,you chose 80 employees and observe the average days absent is 3.6 and the standard deviation is 0.65.

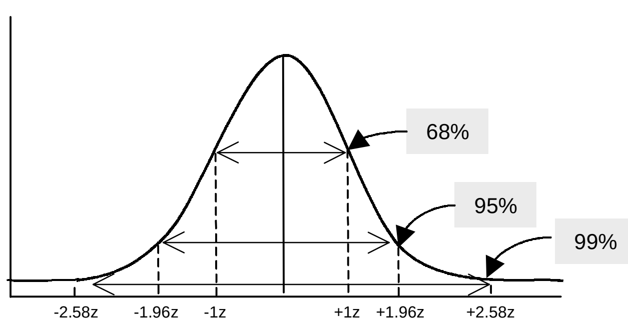
Are you sure that the average number of days absent is representative of the actual days absent for the population? You must consider the chance of committing an error by random sampling. How can you estimate the margin of error?

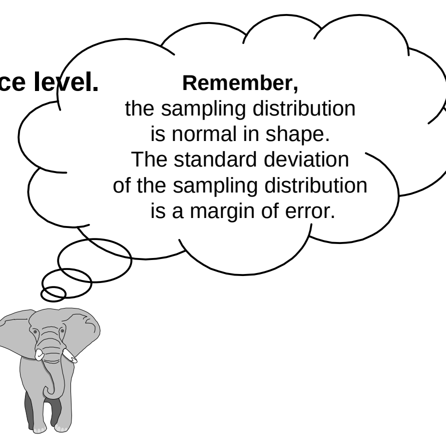
Let's use inferential statistics. You are seeking a 95% confidence level. That is, you want to be confident with a certainty level of 95%.

STEP 1: Identify all of the descriptive statistics information you have.

Sample = 80
Mean = 3.6
Standard Deviation = 0.65
Seeking a 95% confidence level

STEP 2: Identify the z-value for the predefined confidence level.





In this case, the z-value = 1.96 (at 95% confidence level).

STEP 3: Calculate the value of the margin of error.

$$\text{Margin of Error} = \text{z-value}\left(\frac{\text{standard deviation}}{\sqrt{\text{sample number -1}}}\right) = 1.96\left(\frac{0.65}{\sqrt{79}}\right) = 0.14$$

STEP 4: Report the confidence interval by adding and subtracting the margin of error to the mean score.

95 % confident that the population mean will be $3.6 \pm 0.14 = 3.46$ to 3.74

To reduce the margin or error: Use reasonably large samples and use unbiased sampling

GUIDELINES

Point and Interval Estimates

A **point estimate** is a statistic that is computed from a sample. It is an estimate of a population parameter. For example, in surveying a sample of workers in a plant, the workers were asked their age in years. The average (mean) age of the sample was 33 years. An analyst can conclude that 33 years is the best estimate of the average age of the population; that is, the best point estimate of what the average age would be if all of the workers had been surveyed.

A **confidence interval** is a range of values for a sample statistic that is likely to contain the population parameter from which the sample statistic was drawn. The interval will include the population parameter a certain percentage of the time. The percentage chosen is called the **confidence level**. The interval is called the **confidence interval**, its endpoints are **confidence limits**, and the degree of confidence is called the **confidence coefficient**.

Figure 5.5 provides an example of how to calculate a confidence interval.

Returning to the average age of a sample of workers, the point estimate statistic (the average or mean age) was 33. Let's assume we have calculated the estimated standard error of the mean (SEM) and it equals 1.5. We can then say that 95% of the time the population mean will be between:

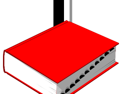
33.0 - (1.96 *1.5) and 33.0 + (1.96*1.5)

or

30.1 years of age and 35.9 years of age

The standard error of the mean (SEM) is calculated using a formula that includes the standard deviation (SD) of the sample and the size of the sample (n). The formula is:

$$\text{SEM} = \frac{\text{SD}}{\sqrt{n-1}}$$



See page 15 in Section 4 for further information about Standard Deviations.

Why plus (+) and minus (-) 1.96? Because 1.96 z equals 95% or (.95) of the total area under the normal curve. The "z" is the symbol for a z-score. A z-score is a simple form of a standard score.

The z-values and formulas for construction of confidence intervals for a mean score for the confidence coefficient of 90%, 95%, and 99% are:

Confidence Coefficient	z-Values	Mean Scores Confidence Intervals Formulas
.90	-1.655 and +1.645	$\bar{X} \pm 1.645 \left(\frac{sd}{\sqrt{n-1}}\right)$
.95	-1.96 and +1.96	$\bar{X} \pm 1.96 \left(\frac{sd}{\sqrt{n-1}}\right)$
.99	-2.575 and +2.575	$\bar{X} \pm 2.575 \left(\frac{sd}{\sqrt{n-1}}\right)$

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

Figure 5.6: Interpreting the t-Test

Since the t-test is used to test the difference between two sample means for statistical significance, report the values of the means, the values of the standard deviations, and the number of cases in each group. The results of a t-test can be described in different ways. Examples include:

Example 1: The difference between the means **is** statistically significant.

($t = 3.01$, $df = 11$, $p < .05$, two-tailed test)

or

The difference between the means **is not** statistically significant.

($t = 1.72$, $df = 13$, $p > .05$, two-tailed test)

Example 2: The difference between the means **is** significant at the .05 level.

($t = 3.01$, $df = 11$, two-tailed test)

or

For the difference between the means **is not** significant at the .05 level.

($t = 1.72$, $df = 13$, *n.s.*, two-tailed test)

NOTE: abbreviation *n.s.* means not significant

Example 3: The null hypothesis was rejected at the .05 level

($t = 3.01$, $df = 11$, two-tailed test)

or

The null hypothesis for the difference between the means was **not** rejected at the .05 level

($t = 1.72$, $df = 13$, two tailed test)

GUIDELINES

THE t-TEST FOR INDEPENDENT SAMPLES:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{\sum X_1^2 - \frac{(\sum X_1)^2}{n_1} + \sum X_2^2 - \frac{(\sum X_2)^2}{n_2}}{(n_1 - 1) + (n_2 - 1)}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$\bar{X}_1$	=	mean score for sample 1
$\bar{X}_2$	=	mean score for sample 2
$\sum X_1^2$	=	sample 1 scores squared and summed
$\sum X_2^2$	=	sample 2 scores squared and summed
$(\sum X_1)^2$	=	sample 1 scores summed and total value squared
$(\sum X_2)^2$	=	sample 2 scores summed and total value squared
n_1	=	number of scores in sample 1
n_2	=	number of scores in sample 2

Figure 5.6 shows how to interpret the t-test.

Definition: A process for determining if there is a statistically significant difference between the means of two independent samples.

Characteristics: Technically called Student's t-Distribution because the author who made the t-test, W.S. Gossett, used the pen name "student." It is useful for interpreting data from small samples when little or nothing is known about the variances of the populations. It also works well with large samples. It is a very robust statistical test.

When to Use: To make inferences about differences between two population means using samples from two populations. The samples must be drawn from two populations or be two independently drawn samples from the same population. Interval or ration data are required.

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

Table 5.2: Portion of a t-Test Table

Critical Values for the t-Distribution

	.05 level of significance		**.01** level of significance	
df	one-tailed test	two-tailed test	one-tailed test	two-tailed test
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947

From: Hinton, P.R. (1996) Statistics Explained: A Guide For Social Science Statistics. New York, N.Y: Rantledge. p.307

GUIDELINES

THE t-TEST FOR DEPENDENT SAMPLES:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\sum d^2 - \dfrac{(\sum d)^2}{n}}{n(n-1)}}}$$

df = n-1

where

$\bar{X}_1$	=	mean for sample 1
$\bar{X}_2$	=	mean for sample 2
d	=	difference in any pair of scores
$\sum d^2$	=	difference in pairs of scores squared and summed
$(\sum d)^2$	=	difference in pairs of scores summed and total value squared
n	=	number of pairs of scores

Definition: A process for determining if there is a statistically significant difference between the means of two dependent samples

When to Use: When all subjects contribute a score to both samples. Such data are dependent or related. Dependent data are obtained when each score or value in one set of data are paired with a score or value in another set.

USING A t-TEST TABLE:

Table 5.2 displays part of the critical value distribution for the t-Test.

Tables with critical values for the t-distribution can be found in most basic statistical textbooks. The numeric values in a t-table indicate the critical values for t at different degrees of freedom for different levels of significance. In order for a t-test value to be significant, the observed value must exceed the critical value for the chosen level of significance.

A word about degrees of freedom. Usually abbreviated “df”, it represents the number of values free to vary when computing a statistic. It is necessary to interpret various inferential statistical tests. The various statistic tests have clear rules for how to calculate df given the number of values free to vary in each sample

NOTES AND SUPPLEMENTAL INFORMATION

GUIDELINES

Figure 5.7: Types of Analysis of Variance (ANOVA)

One-Way ANOVA:

When two or more means are compared on a single variable using ANOVA, this is called a one-way ANOVA. For example, if the absentee rate of salaried workers in three different manufacturing plants is compared and the average (mean) number of days missed for the month of January for the samples for each plant are:

Plant 1: $\bar{X}_1 = 1.78$

Plant 2: $\bar{X}_2 = 3.98$

Plant 3: $\bar{X}_3 = 6.12$

This involves one variable, absentee rate, and calls for a one-way ANOVA.

Two-Way ANOVA:

In a two-way ANOVA (sometimes called a two-factor ANOVA) subjects are classified in two ways. For example, if the absentee rates from above are broken out by gender, i.e., male and female, for plant 1 and 2, there would be four mean scores. For example:

Plant Designation

	#1	#2	Row Means
Female	$\bar{X} = 1.63$	$\bar{X} = 3.01$	$\bar{X} = 2.82$
Male	$\bar{X} = 1.89$	$\bar{X} = 4.21$	$\bar{X} = 3.09$
Column Means	$\bar{X} = 1.98$	$\bar{X} = 3.98$	

This involves two variables, absentee rates and gender, and calls for a two-way ANOVA.

Multivariate Analysis of Variance (MANOVA):

The extension of ANOVA techniques to studies with three or more variables classified in two or more ways is called multivariate analysis of variance. Such statistics require the gathering and analyses of complex sets of data, and it is suggested that a statistical consultant be used.

The mean scores for such a study might look like the following:

Absentee Rates

Shift	Plant #1	Plant #2	Plant #3
1st	$\bar{X} = 1.50$	$\bar{X} = 2.98$	$\bar{X} = 4.98$
2nd	$\bar{X} = 2.01$	$\bar{X} = 3.01$	$\bar{X} = 6.87$
3rd	$\bar{X} = 1.80$	$\bar{X} = 4.89$	$\bar{X} = 7.09$

ANALYSIS OF VARIANCE (ANOVA):

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F
Between groups	SS_B	df_b	$MS_B = SS_B/df_B$	$F = MS_B/MS_w$
Within groups	SS_w	df_w	$MS_w = SS_w/df_w$	
Total	SS_T	df_T		

where

SS_T = total sum of squares
SS_B = between group sum of squares and mean square
$SS_w = SS_T – SS_B$
df_T = number of individual scores –1
df_B = number of groups –1
df_w = number of individual scores – number of groups

Figure 5.7 explains different types of analysis of variance.

Definition: Analysis of variances is used to test for differences between three or more sample mean scores. ANOVA is sometimes called the F-test.

Characteristics: When three or more mean scores are being compared, ANOVA allows the analyst to test the differences between all groups. ANOVA calculates an F- statistic (an F ratio).

At the heart of the variance calculation is the sum of squares (SS): $\sum (X - \bar{X})^2$

It measures the variability of the scores from the mean of the sample. When scores vary wildly from the mean, the sums of the squares will be large. When they cluster around the mean, the sums of squares is small. The sums of square is impacted by the number of scores in a sample, more scores yield larger sums of squares. This is adjusted for by calculating and applying the degrees of freedom.

ANOVA can be used to test the difference between two mean scores. When this is done, the t-test and ANOVA yield the same probability.

When to Use: By calculating the variability in data and producing an F- Test, the ratio of the variances provide an estimate of the statistical difference between the means. If the value of F is greater than the critical value of the F distribution at the chosen level of significance, then the null hypothesis is rejected. It can be concluded that there are statistically significant differences between at least some of the means.

As with all statistical tests, if the F-test is statistically significant, the analyst must decide if the findings are of practical significance.

GUIDELINES

NOTES AND SUPPLEMENTAL INFORMATION

Figure 5.8: Chi-Square Example

- Used in situations where observations can be placed in one of several different (mutually exclusive) categories called cells.
- The data are frequencies or percentages of nominal (categorical) or ordinal (ordered) data.
- In all but the most extreme cases, it is difficult to tell by looking if one set of frequencies differ significantly from another.

1. Present data in a contingency table and calculate row, column, and grand totals.

Example Contingency Table

Age	Agree	Disagree	No Opinion	Row Total
Under 25	5	22	6	33
25 - 40	5	16	6	27
Over 40	30	5	17	40
Column Total	40	43	17	Grand Total 100

2. Calculate expected frequencies. For each cell, multiply the associated row total by the associated column total and divide by the grand total. For example, in cell 1 expected figures are: (33 x 40) / 100 = 13.2. Continue for each cell.

3. Now, apply the Chi-Square formula: $X^2 = \sum \frac{(\bar{O} - E)^2}{E} = 36.0$

4. Compute the degrees of freedom. df = (number of rows –1) (number of columns –1)

$$df = (3-1)(3-1) = 4$$

5. Reference a standard table of Chi-Square values and choose a probability level (i.e., .10, .05, or, .01, etc.)

6. At the .05 level of significance, the Chi-Square table indicates that a value less than or equal to 9.49 is <u>not</u> statistically significant.

7. In the example, $X^2 = 36.0$. One can be certain with 95% probability, that the differences among the groups is due to something other than chance. In short, the sample respondents do have different views on the subject at hand!

GUIDELINES

THE CHI-SQUARE TEST (X^2)

$$X^2 = \sum \frac{(O - E)^2}{E}$$

where

O = observed frequencies

E = expected frequencies

Figure 5.8 is an example of a Chi-Square test.

Definition: Chi-Square is used to test differences among sample frequency (nominal) data.

Characteristics: The most frequent use of the Chi-Square test is to ascertain if there are statistically significant differences between observed frequencies and expected (or hypothetical) frequencies for two or more variables. The frequency data are typically presented in a contingency table or cross-tabulation. The larger the observed discrepancy is in comparison to the expected frequency, the larger the Chi-Square statistic and the more likely the observed differences are statistically significant.

When to Use: The Chi-Square test is used with categorical (nominal or naming) data represented as frequency counts or percentages. The analyst collects frequency data indicating the number of subjects in each category and arranges the data in a table format.

The Chi-Square technique enables the analyst to answer two types of questions. First, an analyst can test how one distribution might differ from a predetermine theoretical distribution. This is called “testing goodness of fit.”

Second, the analyst can compare responses obtained from among or between various groups. This is called “testing for independence.” Both types of situations are tested using the same procedure.

There are two broad types of inferential statistics.

Parametric techniques assume data have certain characteristics: usually they approximate a normal distribution and are measured with interval or ratio scales.

Nonparametric techniques are used when the data depart from the normal distribution and the measures can be nominal or ordinal.

A Final Word on Inferential Statistics

Inferential statistics is the branch of statistics that involves drawing conclusions from sample data. This means drawing an inference about a population on the basis of sample data from the population. One must be clear in identifying a population and explain the rules used for picking a sample.

Inferential statistics tell us how much confidence we can have when we generalize from a sample to a population. In choosing the proper inferential statistic, an analyst must consider the size and number of samples drawn and the nature of the measures taken. When in doubt, it is best to reference a basic textbook on statistics and to seek the advice of an expert.